MIT 18.06 线性代数笔记

Lecturer: Gilbert Strang

Recorder: 陈乐旋

目录

第一讲　线性方程组的几何意义

线性方程组、矩阵

由 n 个方程以及 n 个未知数构成的线性方程组的一般形式：

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n \end{cases}$$

对应的矩阵形式（记为 $\boldsymbol{Ax} = \boldsymbol{b}$）：

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

行图像

所谓的行图像（*row image*）是指一次只考虑 $\boldsymbol{A}$ 中的一行（或方程组中的一个方程），根据该行（方程）在 n 维空间绘制出来的图像称为行图像，比如第 i 个方程 $a_{i1}x_1+a_{i2}x_2+\cdots+a_{in}x_n = b_i$，对应的矩阵形式为：

$$\begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_i \end{bmatrix}$$

列图像

而列图像（*column image*）则是将 $\boldsymbol{A}$ 分割成 n 个列向量绘制在 n 维空间并组合成目标向量 $\boldsymbol{b}$，

$$\begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{bmatrix} x_1 + \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{bmatrix} x_2 + \cdots + \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{nn} \end{bmatrix} x_n = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

记为 $\boldsymbol{a}_1x_1 + \boldsymbol{a}_2x_2 + \cdots + \boldsymbol{a}_nx_n = \boldsymbol{b}$，其中 $\boldsymbol{a}_i$ 为 $\boldsymbol{A}$ 的第 i 列，所以求解原方程组又可理解为：在 n 维空间中，给定 n 个向量（即 A 的 n 个列向量），如何组合出向量 $\boldsymbol{b}$。

示例一

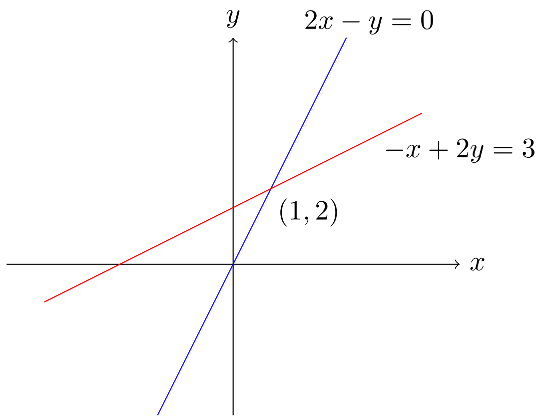
给定如下方程组 (二维)：

$$\begin{cases} 2x - y = 0 \\ -x + 2y = 3 \end{cases}$$

对应的矩阵形式：

$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

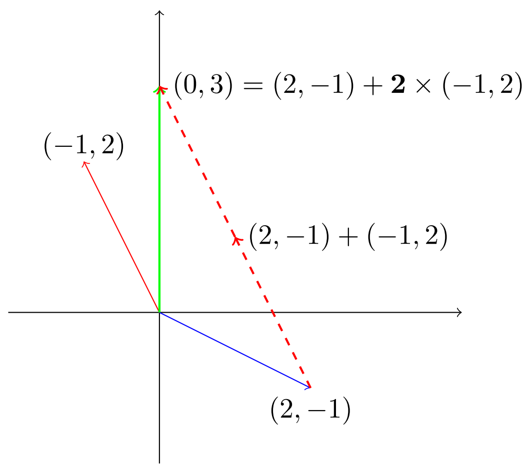
对应的行图像则是将每个方程 (矩阵 $\boldsymbol{A}$ 的每行) 确定的限制条件绘制在二维坐标系中，交点即为方程组的解：



对应的列图像则是用矩阵的列向量组合出目标向量 $\boldsymbol{b}$

$$\begin{bmatrix} 2 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \end{bmatrix} \Rightarrow \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

分别绘制在二维空间中 (蓝、红向量)，并组合出目标向量 (绿色向量)：

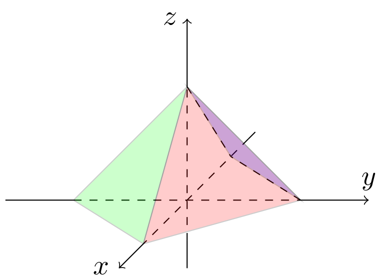


方程组的解即为列向量的组合系数，比如上图中 1 个蓝色向量和 2 个红色向量可组合成绿色向量，所以方程组的解为 $(1, 2)$。(注：列图像的坐标轴没有标注 x, y 轴，因为准确的标注应该为列向量的分量，而不是未知数 x, y)

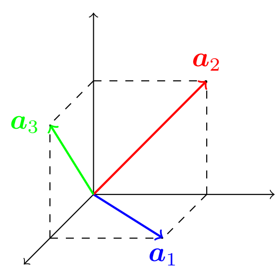
扩展

对于高维情况来说，行、列图像的理解方式可以依次类推，比如对于三维情况下的行图像，每个方程可以确定三维坐标系中的一个平面，对应的解为平面之间的交 (比如点、线、面) 或无解 (即无交点，比如三个平行但不相交的平面)，而三维情况下的列图像则是用三维空间中的三个 (列) 向量组合出方程组右侧的 $\boldsymbol{b}$，对应的解为向量的组合系数，无解则意味着无法用列向量组合出 $\boldsymbol{b}$ (或者说 $\boldsymbol{b}$ 不在 $\boldsymbol{A}$ 的列向量空间中)；对于 n 维情况下的行图像，每个方程可以确定 n 维坐标系中的某个 $n-1$ 维 “平面”，对应的解则为这些 “平面” 之间的交点或无解 (无交点，比如 n 个 “平行” 但不相交的 “平面”)，而 n 维情况下的列图像则是 $\boldsymbol{A}$ 的 n 个列向量 (n 维空间中) 来组合出方程组右侧的向量 $\boldsymbol{b}$ (n 维空间中)。

3-dim row image



3-dim column image



总结

这一讲的重点在于线性方程组、矩阵和几何空间的关系，并且强调了列图像 (***column image***) 的重要性，对于高维矩阵，行图像会变得很复杂，而列图像理解起来则要简单的多。

第二讲　矩阵消元

大纲

消元、回代、消元矩阵

消元

所谓消元 (***Elimination***) 就是通过方程之间的加减来简化方程组 (对应矩阵或增广矩阵的行运算)，从而更容易地求解。

示例一

求解如下方程组：

$$\begin{cases} x+2y+z=2 \\ 3x+8y+z=12 \\ 4y+z=2 \end{cases}$$

对应增广矩阵如下：

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 2 \\ 3 & 8 & 1 & 12 \\ 0 & 4 & 1 & 2 \end{array}\right]$$

消元过程如下：

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 2 \\ 3 & 8 & 1 & 12 \\ 0 & 4 & 1 & 2 \end{array}\right] \xrightarrow{(2,1)} \left[\begin{array}{ccc|c} 1 & 2 & 1 & 2 \\ 0 & 2 & -2 & 6 \\ 0 & 4 & 1 & 2 \end{array}\right] \xrightarrow{(3,2)} \left[\begin{array}{ccc|c} 1 & 2 & 1 & 2 \\ 0 & 2 & -2 & 6 \\ 0 & 0 & 5 & -10 \end{array}\right]$$

箭头上方的内容代表消元目标的位置，比如 $(2,1)$ 代表希望将 a_{21} 或者说矩阵第 2 行、第 1 列的元素消除 (置零)。(注：上述消元的过程应该包括 $(3,1)$，但因为 a_{31} 已经为 0，所以省略了，但 matlab 不会省略这一步；同样，matlab 在消元时 $\boldsymbol{A}$ 和 $\boldsymbol{b}$ 的消元是分开进行的，而不是以增广矩阵的方式一次性完成；最后，当出现主元为 0 的情况时，需要进行一定的行交换。)

其中，矩阵 $\boldsymbol{A}$ 经过消元之后得到的上三角矩阵记为 $\boldsymbol{U}(\boldsymbol{U}pper)$：

$$\boldsymbol{U} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 0 & 5 \end{bmatrix}$$

在高斯消元法中，这种每一行中第一个不为零的元素被称为主元 (***pivot***)，如下图所示：

$$\left[\begin{array}{ccc|c} \boxed{1} & 2 & 1 & 2 \\ 3 & 8 & 1 & 12 \\ 0 & 4 & 1 & 2 \end{array}\right] \xrightarrow{(2,1)} \left[\begin{array}{ccc|c} \boxed{1} & 2 & 1 & 2 \\ 0 & \boxed{2} & -2 & 6 \\ 0 & 4 & 1 & 2 \end{array}\right] \xrightarrow{(3,2)} \left[\begin{array}{ccc|c} \boxed{1} & 2 & 1 & 2 \\ 0 & \boxed{2} & -2 & 6 \\ 0 & 0 & \boxed{5} & -10 \end{array}\right]$$

回代

回代 (***Back-Substitution***) 就是消元之后，从下面的方程解出未知数后依次代入上面的方程求解。

示例一 (续)

示例一中经过消元之后得到了一个更加简洁的矩阵 (对角线以下的元素均为 0)：

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 2 \\ 0 & 2 & -2 & 6 \\ 0 & 0 & 5 & -10 \end{array}\right]$$

对应的方程组为：

$$\begin{cases} x+2y+z=2 \\ 2y-2z=6 \\ 5z=-10 \end{cases}$$

根据第三个方程可以很简单地计算出 $z=-2$，再把 $z=-2$ 代入第二个方程计算出 y，再把 z,y 代入第一个方程计算出 x，这个过程就是回代。

消元矩阵

消元矩阵 (***Elimination Matrix***) 就是表示消元过程的矩阵。

示例一 (续)

示例一中的消元过程 (行变换) 如下：

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 2 \\ 3 & 8 & 1 & 12 \\ 0 & 4 & 1 & 2 \end{array}\right] \xrightarrow{(2,1)} \left[\begin{array}{ccc|c} 1 & 2 & 1 & 2 \\ 0 & 2 & -2 & 6 \\ 0 & 4 & 1 & 2 \end{array}\right] \xrightarrow{(3,2)} \left[\begin{array}{ccc|c} 1 & 2 & 1 & 2 \\ 0 & 2 & -2 & 6 \\ 0 & 0 & 5 & -10 \end{array}\right]$$

因为矩阵的行变换等价于左乘相应的矩阵；矩阵的列变换等价于右乘相应的矩阵。

所以消元过程可表示为左乘一系列的矩阵，比如 $(2,1)$ 的过程是要将 a_{21} 置零，也就是要用第二行减去 3 倍的第一行，用矩阵表示的话就是：

$$(2,1) \Longleftrightarrow \begin{bmatrix} 1 & 0 & 0 \\ \boxed{\text{-3}} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \left[\boldsymbol{A} \,\middle|\, \boldsymbol{b}\right] = \left[\boldsymbol{A}' \,\middle|\, \boldsymbol{b}'\right]$$

记该左乘矩阵为 $\boldsymbol{E}_{21}$；对于 $(3,2)$ 过程采用同样的方式表示，$(3,2)$ 过程是要将 a_{32} 置零，也就是要用第三行减去 2 倍的第二行，则

$$(3,2) \Longleftrightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \boxed{\text{-2}} & 1 \end{bmatrix} \boldsymbol{E}_{21} \left[\boldsymbol{A} \,\middle|\, \boldsymbol{b}\right] = \left[\boldsymbol{A}'' \,\middle|\, \boldsymbol{b}''\right]$$

记该左乘矩阵为 $\boldsymbol{E}_{32}$。

则例一消元过程可表示为：$\boldsymbol{E}_{32}\boldsymbol{E}_{21}\left[\boldsymbol{A}\,\middle|\,\boldsymbol{b}\right]$，且满足 $\boldsymbol{E}_{32}\boldsymbol{E}_{21}\boldsymbol{A}=\boldsymbol{U}$，其中

$$\boldsymbol{E}_{21}=\begin{bmatrix}1&0&0\\-3&1&0\\0&0&1\end{bmatrix},\boldsymbol{E}_{32}=\begin{bmatrix}1&0&0\\0&1&0\\0&-2&1\end{bmatrix}$$

由 $\boldsymbol{U}$ 到 $\boldsymbol{A}$ 则可表示为：$\boldsymbol{E}_{21}^{-1}\boldsymbol{E}_{32}^{-1}\boldsymbol{U}=\boldsymbol{A}$

总结

这一讲比较基础但非常重要，介绍了高斯消元法以及用消元矩阵来表示高斯消元的过程。

第三讲　乘法和逆矩阵

大纲

矩阵乘法的 4 种方法、$\boldsymbol{A}, \boldsymbol{AB}, \boldsymbol{A}^T$ 的逆、高斯-若尔当方法计算 $\boldsymbol{A}^{-1}$

矩阵乘法

矩阵乘法 $\boldsymbol{AB}=\boldsymbol{C}$ 的 4 种计算方法 (理解方式)：

1. 行乘列：单个元素计算，$c_{ij}=\sum_{k=1}^{n} a_{ik}b_{kj}$
2. 行、列：$\boldsymbol{C}$ 的每一列都是 $\boldsymbol{A}$ 的列向量的线性组合；$\boldsymbol{C}$ 的每一行都是 $\boldsymbol{B}$ 的行向量的线性组合
3. 列乘行：$\boldsymbol{C}=\sum_{k=1}^{n}\boldsymbol{a}_k\boldsymbol{b}_k$，其中 $\boldsymbol{a}_k$ 是 $\boldsymbol{A}$ 的第 k 列，$\boldsymbol{b}_k$ 是 $\boldsymbol{B}$ 的第 k 行
4. 分块

$$\begin{bmatrix}\boldsymbol{A}_1 & \boldsymbol{A}_2\\ \boldsymbol{A}_3 & \boldsymbol{A}_4\end{bmatrix}\begin{bmatrix}\boldsymbol{B}_1 & \boldsymbol{B}_2\\ \boldsymbol{B}_3 & \boldsymbol{B}_4\end{bmatrix}=\begin{bmatrix}\boldsymbol{A}_1\boldsymbol{B}_1+\boldsymbol{A}_2\boldsymbol{B}_3 & \boldsymbol{A}_1\boldsymbol{B}_2+\boldsymbol{A}_2\boldsymbol{B}_4\\ \boldsymbol{A}_3\boldsymbol{B}_1+\boldsymbol{A}_4\boldsymbol{B}_3 & \boldsymbol{A}_3\boldsymbol{B}_2+\boldsymbol{A}_4\boldsymbol{B}_4\end{bmatrix}$$

逆矩阵

$\boldsymbol{A}^{-1}\boldsymbol{A}=\boldsymbol{I}=\boldsymbol{A}\boldsymbol{A}^{-1},\ (\boldsymbol{AB})^{-1}=\boldsymbol{B}^{-1}\boldsymbol{A}^{-1},\ (\boldsymbol{A}^T)^{-1}=(\boldsymbol{A}^{-1})^T$

高斯-若尔当

$$\begin{bmatrix}\boldsymbol{A} \mid \boldsymbol{I}\end{bmatrix}\xrightarrow{\text{row transformation}}\begin{bmatrix}\boldsymbol{I} \mid \boldsymbol{A}^{-1}\end{bmatrix}$$

总结

这一讲介绍了从四个角度理解矩阵乘法，角度一是从矩阵 $\boldsymbol{C}$ 的单个元素来看，每个元素都是 $\boldsymbol{A}$ 对应行和 $\boldsymbol{B}$ 的对应列的内积结果，角度二则是从矩阵 $\boldsymbol{C}$ 的单行或单列来看，$\boldsymbol{C}$ 的每一行都是 $\boldsymbol{B}$ 的行向量的线性组合，$\boldsymbol{C}$ 的每一列都是 $\boldsymbol{A}$ 的列向量的线性组合，角度三则是从矩阵 $\boldsymbol{C}$ 的整体来看，矩阵 $\boldsymbol{C}$ 等于 n 个秩为 1 的矩阵相加 (秩为 1 说明行向量或列向量都是对应呈比例的，或者说在线性空间的一条直线上)，角度四则体现了分治的算法思想，比如应用在 *strassen* 算法中。除了矩阵乘法，这一讲还介绍了矩阵逆的概念以及如何使用高斯-若尔当方法来计算逆矩阵。

第四讲 $\boldsymbol{A}=\boldsymbol{L}\boldsymbol{U}$ 分解

大纲

$\boldsymbol{A}=\boldsymbol{L}\boldsymbol{U}$ 分解、置换矩阵

$\boldsymbol{A}=\boldsymbol{L}\boldsymbol{U}$ 分解

第二讲中介绍过了消元矩阵：$\boldsymbol{E}\boldsymbol{A}=\boldsymbol{U}$，其中 $\boldsymbol{E}=\cdots\boldsymbol{E}_{32}\boldsymbol{E}_{31}\boldsymbol{E}_{21}$，$\boldsymbol{E}$ 是由一系列的消元矩阵组合而成的。而 $\boldsymbol{A}=\boldsymbol{L}\boldsymbol{U}$ 分解也就是在上述等式两边分别左乘 $\boldsymbol{E}$ 的逆，$\boldsymbol{A}=\boldsymbol{E}^{-1}\boldsymbol{U}=\boldsymbol{L}\boldsymbol{U}$，其中 $\boldsymbol{L}=\boldsymbol{E}^{-1}=\boldsymbol{E}_{21}^{-1}\boldsymbol{E}_{31}^{-1}\boldsymbol{E}_{32}^{-1}\cdots$($\boldsymbol{L}$ 是下三角矩阵的首字母)。而 $\boldsymbol{A}=\boldsymbol{L}\boldsymbol{U}$ 分解的一个优点在于 (相比 $\boldsymbol{E}\boldsymbol{A}=\boldsymbol{U}$ 的形式) 矩阵 $\boldsymbol{L}$ 非常简洁地记录了消元乘数，而不会引入消元乘数的乘积。

示例一

假设三阶方阵 $\boldsymbol{A}$ 经过下列消元 (行变换) 得到 $\boldsymbol{U}$：$\boldsymbol{E}_{32}\boldsymbol{E}_{31}\boldsymbol{E}_{21}\boldsymbol{A}=\boldsymbol{U}$，其中

$$\boldsymbol{E}_{21}=\begin{bmatrix}1&0&0\\ \boxed{-2}&1&0\\0&0&1\end{bmatrix},\boldsymbol{E}_{31}=\boldsymbol{I},\boldsymbol{E}_{32}=\begin{bmatrix}1&0&0\\0&1&0\\0&\boxed{-5}&1\end{bmatrix},$$

换句话说，$\boldsymbol{A}$ 到 $\boldsymbol{U}$ 的消元过程是先用第二行减去 2 倍的第一行 (消去 a_{21})，而 $\boldsymbol{E}_{31}=\boldsymbol{I}$ 说明 a_{31} 本来就为 0，所以保持不变，再用第三行减去 5 倍的第二行 (消去 a_{32})。用这三个矩阵计算 $\boldsymbol{E}$：

$$\boldsymbol{E}=\boldsymbol{E}_{32}\boldsymbol{E}_{31}\boldsymbol{E}_{21}=\begin{bmatrix}1&0&0\\0&1&0\\0&-5&1\end{bmatrix}\boldsymbol{I}\begin{bmatrix}1&0&0\\-2&1&0\\0&0&1\end{bmatrix}=\begin{bmatrix}1&0&0\\-2&1&0\\ \boxed{10}&-5&1\end{bmatrix}$$

可以发现 $\boldsymbol{E}$ 中出现了一个 10，这是因为消元是从上到下进行的，当进行 $\boldsymbol{E}_{32}$ 操作 (用第三行减去 5 倍的第二行) 时，这时的第二行已经减过了 (2 倍的) 第一行，也就是说最终得到的第三行会受到 (10 倍的) 第一行的影响，所以 $\boldsymbol{E}$ 中出现了消元乘数的乘积。再来看看 $\boldsymbol{L}$：

$$\begin{aligned}\boldsymbol{L}=\boldsymbol{E}_{21}^{-1}\boldsymbol{E}_{31}^{-1}\boldsymbol{E}_{32}^{-1}&=\begin{bmatrix}1&0&0\\-2&1&0\\0&0&1\end{bmatrix}^{-1}\boldsymbol{I}^{-1}\begin{bmatrix}1&0&0\\0&1&0\\0&-5&1\end{bmatrix}^{-1}\\&=\begin{bmatrix}1&0&0\\2&1&0\\0&0&1\end{bmatrix}\boldsymbol{I}\begin{bmatrix}1&0&0\\0&1&0\\0&5&1\end{bmatrix}=\begin{bmatrix}1&0&0\\ \boxed{2}&1&0\\0&\boxed{5}&1\end{bmatrix}\end{aligned}$$

可以看出 $\boldsymbol{L}$ 中只记录了消元乘数，没有引入任何其他内容。这可以理解为从 $\boldsymbol{U}$ 变回 $\boldsymbol{A}$ (从下到上) 只需要用 $\boldsymbol{U}$ 的第三行加上 5 倍的第二行 (和 $\boldsymbol{U}$ 的第一行无关，因为 $\boldsymbol{U}$ 第二行相当于 $\boldsymbol{A}$ 中已经减过第一行的第二行)，再用 $\boldsymbol{U}$ 的第二行加上 2 倍的第一行。

消元的代价

对于一个 $n \times n$ 型矩阵，记一个完整的操作为单个元素先进行乘法再进行减法，则消除第一列 $(\boldsymbol{E}_{21}, \boldsymbol{E}_{31}, \cdots, \boldsymbol{E}_{n1})$ 大约需要 n^2 次操作，其中每个 $\boldsymbol{E}_{i1}$ 需要 n 次操作 (因为一行有 n 个元素，每个元素都需要进行一次操作)，依次类推，消除第二列大约需要 $(n-1)^2$ 次操作等等，所有的操作数之和约为

$$1^2 + 2^2 + \cdots + (n-1)^2 + n^2 = \sum_{i=1}^{n} i^2 \approx \int_0^n x^2 \mathrm{d}x = \frac{1}{3} n^3$$

置换矩阵

置换矩阵 (***Permutation Matrix***) 是指每一行和每一列的元素中有且只有一个 1，其余都是 0 的方阵。对于 n 阶方阵来说，总共有 $n!$ 个不同的置换矩阵，这些置换矩阵构成了一个乘法群 (相乘仍在这些置换矩阵中)。用置换矩阵左乘 (右乘) 一个矩阵等价对这个矩阵进行相应的行交换 (列交换)，所以当消元过程中会出现主元为 0 的情形时，可以先用置换矩阵左乘以使得消元过程中不会出现这种情形，详情见下一讲的 $\boldsymbol{PA} = \boldsymbol{LU}$。

总结

这一讲主要介绍了 $\boldsymbol{A} = \boldsymbol{LU}$ 分解，提供了一个看待矩阵消元的不同视角，同时强调了 $\boldsymbol{L}$ 相比 $\boldsymbol{E}$ 的优点；然后简单介绍了置换矩阵 (下一讲的内容)。

第五讲 转置、置换、$\mathbb{R}^n$ 空间

大纲

置换 (续)、$\boldsymbol{PA}=\boldsymbol{LU}$、转置、对称矩阵、向量空间和子空间

置换矩阵 (续)

置换矩阵 (*Permutation Matrix*) 除了上一讲中的内容，还有两个特点：

1. 置换矩阵 $\boldsymbol{P}$ 都是可逆的
2. $\boldsymbol{P}\boldsymbol{P}^T=\boldsymbol{I}, \boldsymbol{P}^{-1}=\boldsymbol{P}^T$

$\boldsymbol{PA}=\boldsymbol{LU}$

第四讲中介绍了 $\boldsymbol{A}=\boldsymbol{LU}$ 分解，并且在第二讲中介绍了矩阵消元 $\boldsymbol{EA}=\boldsymbol{U}$，但这两种形式都没有考虑消元过程中出现主元 (*pivot*) 为 0 并需要交换行的情形，所以完整的形式应该为 $\boldsymbol{PA}=\boldsymbol{LU}$，即先将 $\boldsymbol{A}$ 进行行交换，以使得消元过程中不会出现主元为 0 的情况。

转置矩阵

$$\left[\boldsymbol{A}^T\right]_{ij}=\left[\boldsymbol{A}\right]_{ji}$$

对称矩阵

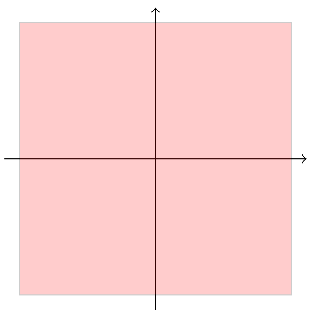
$\boldsymbol{A}^T=\boldsymbol{A}$；任意矩阵 $\boldsymbol{A}$ 乘其转置 $\boldsymbol{A}^T$ 都是对称矩阵：$(\boldsymbol{A}\boldsymbol{A}^T)^T=\boldsymbol{A}\boldsymbol{A}^T$

向量空间

向量空间 (*Vector Space*) 是指由向量构成的集合 (空间)，并且该集合中的元素相加和标量乘法 (线性组合) 依旧在这个集合 (空间) 中，或者说向量空间必须满足线性组合的封闭性。

示例一

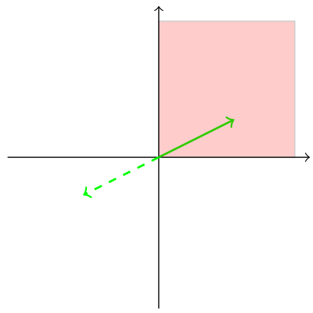
比如所有的二维 (实) 向量构成的空间 $\mathbb{R}^2$ 就是一个向量空间：



这个空间中的任意向量的线性组合 (相加和标量数乘) 都依旧在这个空间中。

示例二

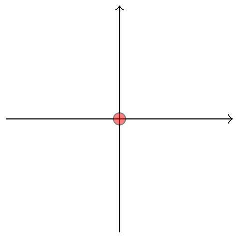
比如所有的分量均为非负数的二维 (实) 向量构成的空间 (类似于第一象限) 就不是一个向量空间：



很明显，这个空间的非零向量标量乘 -1 的话就不在这个空间中了。

示例三

很明显由单个零向量构成的空间是一个向量空间：



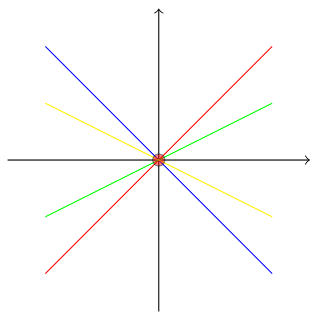
并且所有的向量空间都应该包含对应的零向量 (因为数乘 0 必须在向量空间中)。

子空间

子空间 (***Subspace***) 是指由向量空间中的一部分构成的集合 (或者说空间)，并且该子集合中的元素依旧构成一个向量空间。

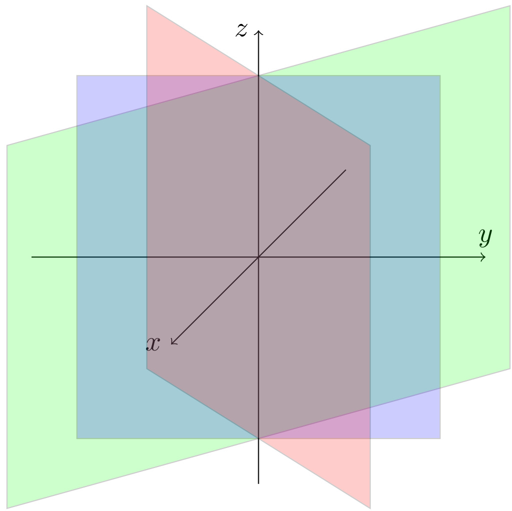
示例四

$\mathbb{R}^2$ 的子空间包括：由单个零向量构成的空间、过零点的任意一条直线上的所有向量构成的空间、$\mathbb{R}^2$ 本身



示例五

$\mathbb{R}^3$ 的子空间包括：由单个零向量构成的空间、过零点的任意一条直线上的所有向量构成的空间、过零点的任意一个平面上的所有向量构成的空间、$\mathbb{R}^3$ 本身



列空间

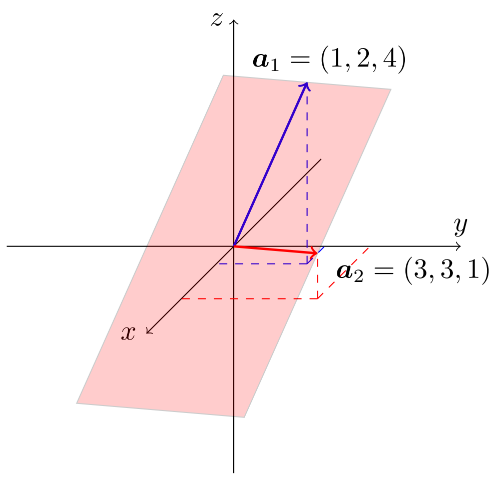
矩阵 $\boldsymbol{A}$ 的列空间（***Column Space***）是指由矩阵的列向量的线性组合构造出的向量空间，记为 $\boldsymbol{C}(\boldsymbol{A})$。

示例六

假设

$$\boldsymbol{A}=\begin{bmatrix}1&3\\2&3\\4&1\end{bmatrix}$$

则对应的列空间 $\boldsymbol{C}(\boldsymbol{A})$ 如下图所示：



总结

这一讲终于开始了对向量空间、子空间的讨论，用 Gilbert Stang 的话来说：“*the beginning of linear algebra*”，并简单介绍了一下列空间（下一讲的重点之一）。

第六讲　列空间和零空间

大纲

向量空间的并集和交集、列空间、零空间

向量空间的并和交

两个子空间的并集不一定是向量空间；两个子空间的交集是向量空间。

列空间

矩阵的列空间 (*Column Space*) 在上一讲已经简单介绍过，这里有两个点需要注意：

1. 对于 $m \times n$ 型矩阵，列空间是 $\mathbb{R}^m$ 的子空间
2. $\boldsymbol{Ax} = \boldsymbol{b}$ 有解等价于 $\boldsymbol{b}$ 在 $\boldsymbol{A}$ 的列空间中；反之，不在

零空间

矩阵 $\boldsymbol{A}$ 的零空间 (*nullspace*) 是指满足 $\boldsymbol{Ax} = \boldsymbol{0}$ 的所有 $\boldsymbol{x}$ 构成的空间 (注：零空间的翻译容易与由单个零向量构成的空间混淆)，记为 $\boldsymbol{N}(\boldsymbol{A})$，对于 $m \times n$ 型矩阵，零空间是 $\mathbb{R}^n$ 的子空间。而对于满足 $\boldsymbol{Ax} = \boldsymbol{b}(\boldsymbol{b} \neq \boldsymbol{0})$ 的所有 $\boldsymbol{x}$ 并不会构成一个向量空间，因为很明显零向量不在其中。

总结

这一讲的重点是列空间和零空间的概念，以及这两个向量空间的基本特点：对于 $m \times n$ 型矩阵，列空间是 $\mathbb{R}^m$ 的子空间，而零空间是 $\mathbb{R}^n$ 的子空间，之后还会继续讨论。

第七讲 $\boldsymbol{Ax}=\mathbf{0}$ 求解：主变量、特解

大纲

行阶梯形矩阵、简化行阶梯形矩阵、主变量、自由变量、特解、计算零空间 ($\boldsymbol{Ax}=\mathbf{0}$)

行阶梯形矩阵

行阶梯形矩阵 (***Row Echelon Form***)：

- 所有非零行 (矩阵的行至少有一个非零元素) 在所有全零行的上面。即全零行都在矩阵的底部。
- 非零行的首项系数，也称作主元，即最左边的首个非零元素，严格地比上面行的主元更靠右。
- 主元所在列，在该元下面的元素都是零 (前两条的推论)。

简化行阶梯形矩阵

简化行阶梯形矩阵 (***Reduced Row Echelon Form, rref***)，又称作行规范形 (***Row Canonical Form***)，记为 $\boldsymbol{R}=rref(\boldsymbol{A})$，满足额外的条件：每个主元是 1，且是其所在列的唯一的非零元素。

主变量

主变量 (***pivot variable***)，即主元所在的列对应的变量，其中主元所在的列又称作主列 (***pivot column***)，主元的个数等于矩阵的秩 r。

自由变量

自由变量 (***free variable***)，即非主元所在的列对应的变量，其中非主元所在的列又称作自由列 (***free column***)，自由变量的个数等于 $n-r$。

特解

特解 (***Special Solution***)，即满足方程 $\boldsymbol{Ax}=\mathbf{0}$ 的特定的解，每一个自由变量会对应一个特解 (更准确地说是线性无关的特解)。

零空间的计算

计算零空间 (求解 $\boldsymbol{Ax}=\mathbf{0}$) 的算法：

1. 消元 (行阶梯形或简化行阶梯形)
2. 确定主变量、自由变量，依次给自由变量赋 1，其余自由变量赋 0，求解每个特解
3. 特解的线性组合即为零空间 (或方程的通解)

补充

当对 $\boldsymbol{A}$ 进行消元得到 $\boldsymbol{R}$ 时，如果主列正好都是前面的列，$\boldsymbol{R}$ 可记为：

$$\boldsymbol{R}=\begin{bmatrix}\boldsymbol{I}_{r\times r} & \boldsymbol{F}_{r\times(n-r)}\\ \mathbf{0}_{(m-r)\times r} & \mathbf{0}_{(m-r)\times(n-r)}\end{bmatrix}=\begin{bmatrix}\boldsymbol{I} & \boldsymbol{F}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}$$

当依次给自由变量赋 1，其余自由变量赋 0，会得到零空间矩阵 (*nullspace matrix*)，记为 $\boldsymbol{N}$，该矩阵的列向量即为 $\boldsymbol{Ax}=\mathbf{0}$ 的线性无关的特解，该矩阵的列空间即为 $\boldsymbol{A}$ 的零空间：

$$\boldsymbol{N}=\begin{bmatrix}-\boldsymbol{F}_{r\times(n-r)}\\ \boldsymbol{I}_{(n-r)\times(n-r)}\end{bmatrix}=\begin{bmatrix}-\boldsymbol{F}\\ \boldsymbol{I}\end{bmatrix}$$

这里 $-\boldsymbol{F}$ 有 r 行，对应 r 个主变量；$\boldsymbol{I}$ 有 $n-r$ 行，对应 $n-r$ 个自由变量 (并且依次赋 1)。

如果主列并不都是前面的列时 (或者说存在自由列在主列前面时)，只需要对 $\boldsymbol{R}$ 进行相应的列交换 (即交换未知数的次序) 就可以得到上述形式，再对 $\boldsymbol{N}$ 进行相应的行交换依旧可以得到零空间矩阵。

总结

这一讲的内容是求解 $\boldsymbol{Ax}=\mathbf{0}$ 或者说计算 $\boldsymbol{A}$ 的零空间，并解释了如何根据 $\boldsymbol{R}$ 的结构来计算零空间矩阵 $\boldsymbol{N}$。

第八讲 $\boldsymbol{Ax}=\boldsymbol{b}$ 求解：行约简形 $\boldsymbol{R}$

大纲

满秩、$\boldsymbol{Ax}=\boldsymbol{b}$ 求解

满秩

对于 $m\times n$ 型矩阵 $\boldsymbol{A}$，行满秩 (*row full rank*) 即 $r=m$，列满秩 (*column full rank*) 即 $r=n$，行列均满秩即 $r=m=n$。

$\boldsymbol{Ax}=\boldsymbol{b}$ 的几何空间理解

对于 $\boldsymbol{Ax}=\boldsymbol{b}$ 是否有解可以有很多种理解方式 (或者说视角)，比如方程组的视角，即是否存在 $x_1,x_2,\cdots,x_n$ 满足方程组；线性表示的视角，即 $\boldsymbol{b}$ 是否可以由 $\boldsymbol{A}$ 的列组成的向量组线性表出；矩阵秩的视角，即是否 $r(\boldsymbol{A})=r(\boldsymbol{A}|\boldsymbol{b})$；等等。

这里主要记录一下从几何空间的视角如何来看待 $\boldsymbol{Ax}=\boldsymbol{b}$。首先看 $\boldsymbol{A}$，假设 $\boldsymbol{A}$ 是一个 $m\times n$ 型矩阵，则要研究的几何空间包括

- $\mathbb{R}^m$ 空间
- $\mathbb{R}^m$ 空间中的子空间：$\boldsymbol{A}$ 的列空间 $\boldsymbol{C}(\boldsymbol{A})$
- $\mathbb{R}^n$ 空间
- $\mathbb{R}^n$ 空间中的子空间：$\boldsymbol{A}$ 的零空间 $\boldsymbol{N}(\boldsymbol{A})$

现在分情况来讨论一下几何空间中发生的事情：

- 行满秩但列不满秩 $(r=m<n)$：$r=m$ 说明了 $\boldsymbol{A}$ 的列空间 $\boldsymbol{C}(\boldsymbol{A})$ 充满了整个 $\mathbb{R}^m$ 空间，所以毫无疑问，$\boldsymbol{b}$ 也一定在这个空间中 (因为 $\boldsymbol{b}$ 是 $\mathbb{R}^m$ 空间中的一个向量)，即 $\boldsymbol{Ax}=\boldsymbol{b}$ 一定有解；而 $n>r$ 则说明了 $\boldsymbol{A}$ 的零空间 $\boldsymbol{N}(\boldsymbol{A})$ ($\mathbb{R}^n$ 空间中的一个 $n-r$ 维子空间) 有无数多个向量，且 $\boldsymbol{Ax}=\boldsymbol{b}$ 的一个特解向量加上 $\boldsymbol{N}(\boldsymbol{A})$ 中的任意一个向量依旧是 $\boldsymbol{Ax}=\boldsymbol{b}$ 的解，所以一定有无数多个解
- 列满秩但行不满秩 $(r=n<m)$：$r<m$ 说明 $\boldsymbol{C}(\boldsymbol{A})$ 并没有充满 $\mathbb{R}^m$，所以对于 $\mathbb{R}^m$ 空间中的一个任意的向量 $\boldsymbol{b}$，它可能在 $\boldsymbol{C}(\boldsymbol{A})$ 中，又可能不在，所以这种情况下既可能有解，也可能无解，如果 $\boldsymbol{b}$ 正好落在了 $\boldsymbol{C}(\boldsymbol{A})$ 中 (即 $r(\boldsymbol{A})=r(\boldsymbol{A}|\boldsymbol{b})$)，则有解，如果 $\boldsymbol{b}$ 没有落在了 $\boldsymbol{C}(\boldsymbol{A})$ 中 (即 $r(\boldsymbol{A})=r(\boldsymbol{A}|\boldsymbol{b})+1$)，则无解；而 $r=n$ 说明了 $\boldsymbol{A}$ 的零空间 $\boldsymbol{N}(\boldsymbol{A})$ 中只有单个零向量，所以如果有解的话解一定是唯一的
- 行满秩且列满秩 $(r=m=n)$：这种情况就简单了，$\boldsymbol{A}$ 的列空间 $\boldsymbol{C}(\boldsymbol{A})$ 充满了整个 $\mathbb{R}^m$，所以一定有解，而且 $\boldsymbol{A}$ 的零空间 $\boldsymbol{N}(\boldsymbol{A})$ 中只有单个零向量，所以一定有唯一解
- 即不是行满秩也不是列满秩 $(r<m,r<n)$：这种情况就是 $\boldsymbol{C}(\boldsymbol{A})$ 并没有充满 $\mathbb{R}^m$，所以可能无解，也可能有解；而又因为 $\boldsymbol{N}(\boldsymbol{A})$ ($\mathbb{R}^n$ 空间中的一个 $n-r$ 维子空间) 中有无数多个向量，所以有解的话一定有无数多个解

$\boldsymbol{R}$ 的形式

再分情况来讨论一下简化行阶梯型 $\boldsymbol{R}$ 的形式 (及解的情况):

$r=m<n$	$r=n<m$	$r=m=n$	$r<m, r<n$
$\boldsymbol{R}=\begin{bmatrix}\boldsymbol{I} & \boldsymbol{F}\end{bmatrix}$	$\boldsymbol{R}=\begin{bmatrix}\boldsymbol{I}\\ \boldsymbol{0}\end{bmatrix}$	$\boldsymbol{R}=\begin{bmatrix}\boldsymbol{I}\end{bmatrix}$	$\boldsymbol{R}=\begin{bmatrix}\boldsymbol{I} & \boldsymbol{F}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}$
∞	0 or 1	1	0 or ∞

对于 $\boldsymbol{R}$ 来说，如果底部没有零行 (行满秩)，则 $\boldsymbol{C}(\boldsymbol{A})$ 一定充满了整个 $\mathbb{R}^m$，所以一定有解；而底部有零行则说明 $\boldsymbol{C}(\boldsymbol{A})$ 没有充满整个 $\mathbb{R}^m$，所以可能无解，也可能有解。而在有解的前提下是有唯一解还是有无数多个解，则要看是否有 $\boldsymbol{F}$，有 $\boldsymbol{F}$ 则有无数多个解，如果没有则只有唯一解。($\boldsymbol{F}$ 列数对应 $\boldsymbol{N}(\boldsymbol{A})$ 的维数、自由变量的个数)

- 0 个解 $\Rightarrow$ $\boldsymbol{R}$ 底部必有零行，$\boldsymbol{C}(\boldsymbol{A})$ 必没有充满了 $\mathbb{R}^m$
- 1 个解 $\Rightarrow$ $\boldsymbol{R}$ 中必没有 $\boldsymbol{F}$，$\boldsymbol{N}(\boldsymbol{A})$ 中必定只有零向量
- ∞ 个解 $\Rightarrow$ $\boldsymbol{R}$ 中必有 $\boldsymbol{F}$，$\boldsymbol{N}(\boldsymbol{A})$ 中必有无数多个向量

$\boldsymbol{Ax}=\boldsymbol{b}$ 有解条件

1. $r(\boldsymbol{A})=r(\boldsymbol{A}|\boldsymbol{b})$：有解 ($\boldsymbol{b}$ 在 $\boldsymbol{A}$ 的列空间中)
   - $r(\boldsymbol{A})=n$，有唯一解 ($\boldsymbol{A}$ 的零空间只有一个零向量)
   - $r(\boldsymbol{A})<n$，有无数多解 ($\boldsymbol{A}$ 的零空间中有无数多个向量)
2. $r(\boldsymbol{A})\neq r(\boldsymbol{A}|\boldsymbol{b})$：无解 ($\boldsymbol{b}$ 不在 $\boldsymbol{A}$ 的列空间中)

$r(\boldsymbol{A})=r(\boldsymbol{A}|\boldsymbol{b})$

对于 $r(\boldsymbol{A})=r(\boldsymbol{A}|\boldsymbol{b})$ 的一种理解是：如果 $\boldsymbol{A}$ 的行向量的某种线性组合得到了零向量，那 $\boldsymbol{b}$ 对应的分量之间相同的线性组合一定可以得到 0。

$\boldsymbol{Ax}=\boldsymbol{b}$ 求解

求解 $\boldsymbol{Ax}=\boldsymbol{b}$ 的算法：

1. 求解 $\boldsymbol{Ax}=\boldsymbol{0}$ 的通解 (上一讲的内容)
2. 求解 $\boldsymbol{Ax}=\boldsymbol{b}$ 的一个特解：消元、消元后所有自由变量赋 0、求解主变量
3. $\boldsymbol{Ax}=\boldsymbol{0}$ 的通解加上 $\boldsymbol{Ax}=\boldsymbol{b}$ 的一个特解即为 $\boldsymbol{Ax}=\boldsymbol{b}$ 的通解

总结

这一讲的内容主要是求解 $\boldsymbol{Ax}=\boldsymbol{b}$ 以及从不同的视角来理解 $\boldsymbol{Ax}=\boldsymbol{b}$

第九讲　线性相关性、基、维数

大纲	线性相关、线性无关、基、维数
线性相关	如果存在不全为 0 的系数 $a_1, a_2, \cdots, a_n$ 使得 $a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \cdots a_n\boldsymbol{v}_n = \boldsymbol{0}$，则称向量组 $\boldsymbol{v}_1, \boldsymbol{v}_2 \cdots, \boldsymbol{v}_n$ 线性相关 (***linearly dependent***)。
线性无关	反之，如果不存在不全为 0 的系数 $a_1, a_2, \cdots, a_n$ 使得 $a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \cdots a_n\boldsymbol{v}_n = \boldsymbol{0}$，则称向量组 $\boldsymbol{v}_1, \boldsymbol{v}_2 \cdots, \boldsymbol{v}_n$ 线性无关 (***linearly independent***)。
基	向量空间 $\boldsymbol{V}$ 的基 (***basis***) 指 $\boldsymbol{V}$ 中的一组向量 (这些向量被称为基向量)，$\boldsymbol{V}$ 中的任意一个向量都可以唯一地表示成基向量的线性组合。
维数	向量空间 $\boldsymbol{V}$ 的维数 (***dimension***) 指基向量的个数。(一个向量空间 $\boldsymbol{V}$ 可以有很多不同的基，但维数一定相等。)

第十讲　四个基本子空间

大纲

四个基本子空间及其特点、矩阵空间

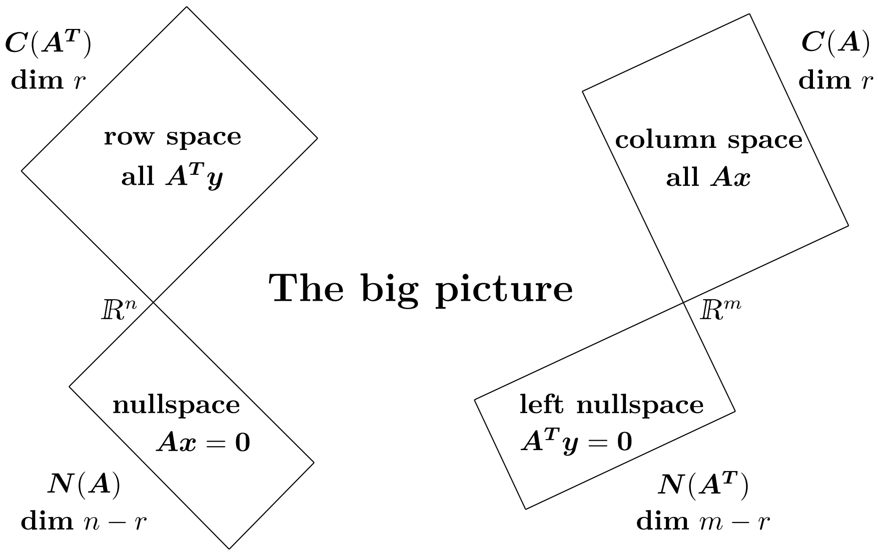
四个基本子空间

矩阵 $\boldsymbol{A}_{m\times n}$ 的四个基本子空间包括：

- 列空间 (*column space*)，记为 $\boldsymbol{C}(\boldsymbol{A})$，$\mathbb{R}^m$ 中的 r 维子空间
- 零空间 (*nullspace*)，记为 $\boldsymbol{N}(\boldsymbol{A})$，$\mathbb{R}^n$ 中的 $n-r$ 维子空间
- 行空间 (*row space*)，记为 $\boldsymbol{C}(\boldsymbol{A^T})$，$\mathbb{R}^n$ 中的 r 维子空间
- 左零空间 (*left nullspace*)，记为 $\boldsymbol{N}(\boldsymbol{A^T})$，$\mathbb{R}^m$ 中的 $m-r$ 维子空间

其中列空间 $\boldsymbol{C}(\boldsymbol{A})$ 和零空间 $\boldsymbol{N}(\boldsymbol{A})$ 在第六讲中已经讲过，而行空间 $\boldsymbol{C}(\boldsymbol{A^T})$，从形式上就可以看出，等价于转置矩阵 $\boldsymbol{A^T}$ 的列空间，左零空间 $\boldsymbol{N}(\boldsymbol{A^T})$ 则是指满足 $\boldsymbol{A^T y}=\boldsymbol{0}$ 的所有解构成的空间 (之所以叫做左零空间，可以将 $\boldsymbol{A^T y}=\boldsymbol{0}$ 转置得 $\boldsymbol{y^T A}=\boldsymbol{0^T}$，这时 $\boldsymbol{y^T}$ 就在 $\boldsymbol{A}$ 的左边了)。

四者关系的宏观视角如下图所示：



求四个基本子空间的基 (注意从线性无关的角度去理解)：

- 列空间：$\boldsymbol{A}$ 的 r 个主列 (*pivot column*) 为基
- 零空间：$\boldsymbol{A}$ 的 $n-r$ 个自由变量依次赋 1、其余均赋 0 所求得的 $n-r$ 个特解为基
- 行空间：$\boldsymbol{A}$ 的简化行阶梯形 $\boldsymbol{R}$ 的 r 个非零行为基
- 左零空间：高斯消元可表示为 $\boldsymbol{EA}=\boldsymbol{R}$，其中 $\boldsymbol{E}$ 为消元矩阵，$\boldsymbol{R}$ 中 $m-r$ 个零行对应的 $\boldsymbol{E}$ 中的 $m-r$ 个行向量即为基

矩阵空间

矩阵空间 (***Matrix Space***) 即由矩阵构成的空间，比如所有的 3×3 型矩阵构成的空间，它的子空间有比如所有的上三角矩阵 (或对称矩阵，或对角矩阵) 构成的的空间等等。

总结

这一讲主要是介绍四个基本子空间，这是线性代数中非常重要的内容，用 Gilbert Stang 的话来说就是 “理解了这四个基本子空间就理解了线性代数一半的内容”；除此之外简单地介绍了一下矩阵空间 (下一讲的内容)。

第十一讲　矩阵空间、秩为 1 的矩阵、小世界图

大纲

矩阵空间、微分方程的解空间、秩为 1 的矩阵

矩阵空间

上一讲结束的时候简单介绍过了矩阵空间 (*Matrix Space*) 的概念，即由矩阵构成的空间 (相比由向量构成的向量空间 (*Vector Space*))。

示例一

比如所有的 3×3 型矩阵构成的矩阵空间 $\boldsymbol{M}$，其中

- 所有的 3×3 型对称矩阵构成的矩阵空间 $\boldsymbol{S}$
- 所有的 3×3 型上三角矩阵构成的矩阵空间 $\boldsymbol{U}$
- 所有的 3×3 型对角矩阵构成的矩阵空间 $\boldsymbol{D}$，即 $\boldsymbol{S}\cap\boldsymbol{U}$

是 $\boldsymbol{M}$ 的三个个子空间。$\boldsymbol{M}$ 的维数是 9，3×3 型矩阵中的每个元素依次取 1、其余元素均 0 所得的 9 个矩阵即为 $\boldsymbol{M}$ 的基；$\boldsymbol{S}$ 的维数 6，如下 6 个矩阵即为 $\boldsymbol{S}$ 的基

$$\begin{bmatrix}1&0&0\\0&0&0\\0&0&0\end{bmatrix},\begin{bmatrix}0&0&0\\0&1&0\\0&0&0\end{bmatrix},\begin{bmatrix}0&0&0\\0&0&0\\0&0&1\end{bmatrix},\begin{bmatrix}0&1&0\\1&0&0\\0&0&0\end{bmatrix},\begin{bmatrix}0&0&1\\0&0&0\\1&0&0\end{bmatrix},\begin{bmatrix}0&0&0\\0&0&1\\0&1&0\end{bmatrix}$$

$\boldsymbol{U}$ 的维数也为 6，如下 6 个矩阵即为 $\boldsymbol{U}$ 的基

$$\begin{bmatrix}1&0&0\\0&0&0\\0&0&0\end{bmatrix},\begin{bmatrix}0&1&0\\0&0&0\\0&0&0\end{bmatrix},\begin{bmatrix}0&0&1\\0&0&0\\0&0&0\end{bmatrix},\begin{bmatrix}0&0&0\\0&1&0\\0&0&0\end{bmatrix},\begin{bmatrix}0&0&0\\0&0&1\\0&0&0\end{bmatrix},\begin{bmatrix}0&0&0\\0&0&0\\0&0&1\end{bmatrix}$$

$\boldsymbol{D}$ 的维数为 3，如下 3 个矩阵即为 $\boldsymbol{D}$ 的基

$$\begin{bmatrix}1&0&0\\0&0&0\\0&0&0\end{bmatrix},\begin{bmatrix}0&0&0\\0&1&0\\0&0&0\end{bmatrix},\begin{bmatrix}0&0&0\\0&0&0\\0&0&1\end{bmatrix}$$

$\boldsymbol{S}+\boldsymbol{U}$(注：不是 $\boldsymbol{S}$ 和 $\boldsymbol{U}$ 的并集) 是两个空间中的任意元素 (矩阵) 的线性组合构成的空间，实际上就是 $\boldsymbol{M}$，所以可得

$$dim(\boldsymbol{S})+dim(\boldsymbol{U})=dim(\boldsymbol{D})+dim(\boldsymbol{M})=dim(\boldsymbol{S}\cap\boldsymbol{U})+dim(\boldsymbol{S}+\boldsymbol{U})$$

解空间

比如如下微分方程：

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2}+y=0$$

实数范围内有两个特解 $y=\sin x$ 和 $y=\cos x$，所有的解即为这两个特解的线性组合：$y=c_1\sin x+c_2\cos x$，所有的解即构成了解空间，解空间的元素为方程的解 (而不是向量、矩阵)，上述两个特解即为该解空间的基，解空间的维数为 2。

秩为 1 的矩阵

秩为 1 的矩阵可以表示为 1 列乘 1 行的形式。秩为 1 的矩阵一个特点在于可以用于构建其他的矩阵，就像搭积木一样，比如秩为 4 的矩阵可以通过 4 个秩为 1 的同型矩阵构建出来。

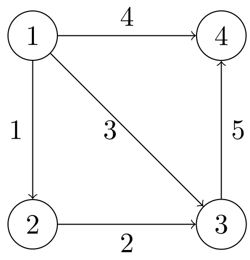
总结

这一讲介绍了矩阵空间和解空间，再加上向量空间，都可以视为线性空间的实例。课程的结束部分简单地介绍了图、图的矩阵表示、六度分割理论 (***Six Degrees of Separation***)。

第十二讲 图、网络、关联矩阵

示例一

比如如下的有向图 (***directed graph***)，有 4 个节点和 5 条边：



用矩阵表示：

$$A = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix},$$

其中每一行对应一条边，每一列对应一个节点，-1 表示从对应的节点出，1 表示从对应的节点入，矩阵 $\boldsymbol{A}$ 的秩为 3。

假设节点 i 的电势为 x_i，记 $\boldsymbol{x} = [x_1, x_2, x_3, x_4]^T$，则 $\boldsymbol{Ax}$ 表示 5 条边所连接的节点之间的电势差，

$$\boldsymbol{Ax} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_2 - x_1 \\ x_3 - x_2 \\ x_3 - x_1 \\ x_4 - x_1 \\ x_4 - x_3 \end{bmatrix}$$

假设我们想求解 5 条边所连接的节点之间的电势差为 0 时各节点的电势，则只需要求解 $\boldsymbol{Ax} = \boldsymbol{0}$，可求得 (零空间)

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = c \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

也就是说，各节点电势差为 0 时，各节点的电势一定相等 (电路的基本原理)。

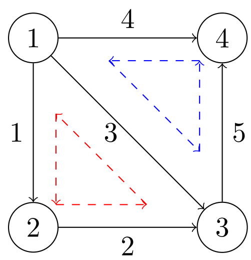
再假设边 i 上的电流为 y_i，记 $\boldsymbol{y} = [y_1, y_2, y_3, y_4, y_5]^T$，则欧姆定律可表示为 $\boldsymbol{y} = \boldsymbol{Ax}/R = C\boldsymbol{Ax}$，其中 $R = 1/C$ 为各边的电阻；基尔霍夫电流定律可表示为 $\boldsymbol{A^T y} = \boldsymbol{0}$，即

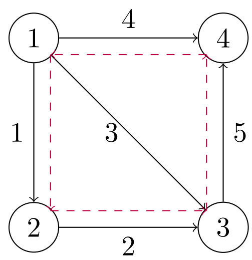
$$\boldsymbol{A^T y} = \begin{bmatrix} -1 & 0 & -1 & -1 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} -y_1 - y_3 - y_4 \\ y_1 - y_2 \\ y_2 + y_3 - y_5 \\ y_4 + y_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

再求解 $\boldsymbol{A^T y}=\mathbf{0}$，可得到如下两个线性无关的解 ($\boldsymbol{N(A^T)}$ 的基)：

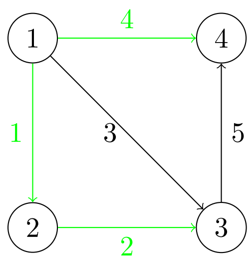
$$\begin{bmatrix}1\\1\\-1\\0\\0\end{bmatrix},\begin{bmatrix}0\\0\\1\\-1\\1\end{bmatrix}$$

对应图中的两个回路 (红、蓝虚线)，若将这两个解相加，则可得解 $[1,1,0,-1,1]^T$，对应图中的较大的回路 (紫色虚线)，而这些解说明了 $\boldsymbol{A^T}$ 中某些列向量的线性相关性，比如第一个解 $[1,1,-1,0,0]^T$ 说明了 $\boldsymbol{A^T}$ 的第 1、2、3 列向量线性相关：





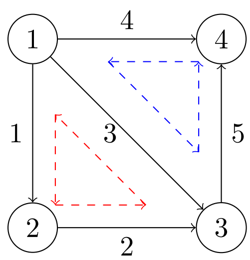
而如果选取 $\boldsymbol{A^T}$ 中线性无关的列向量，比如第 1、2、4 个列向量，在图中对应的则不是回路：



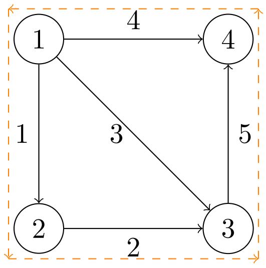
可以看出 $\boldsymbol{A^T}$ 中线性相关的列向量等价于回路，线性无关则不是回路，而且回路的个数等于 $\boldsymbol{N(A^T)}$ 的维数 (注：上图中的紫色回路并不计入，因为可以由红、蓝回路组合得出)。
根据 $\dim(\boldsymbol{N(A^T)})=m-r$，可得 $\#loops=\#edges-(\#nodes-1)$，即 $\#loops+\#nodes-\#edges=1$(图的欧拉公式的一种形式)。

扩展

欧拉公式实际上是 $F+V-E=2$，而不是上式中的 1，这是由于回路的定义问题，在该课程中，计算回路数量时只包括了图内部的回路，也就是



但是并没有包括外部的回路 (注：该回路并不是前面的紫色回路，紫色回路是内部区域，而该回路是朝无限远的地方，也可以想象该图是画在球面上的，该回路是朝球面的对面延伸的)：



总结

这一讲主要是讲线性代数的应用。

第十三讲 测验一复习

例一

设 $\boldsymbol{u}$、$\boldsymbol{v}$、$\boldsymbol{w}$ 是 $\mathbb{R}^7$ 中的非零向量，由它们生成的 $\mathbb{R}^7$ 中的子空间的维数可能是多少？
答：1、2、3

例二

5×3 型的矩阵 $\boldsymbol{U}$ 秩为 3，求 $\boldsymbol{U}$ 的零空间？
答：由单个零向量构成的 0 维空间

例三

10×3 型的矩阵 $\boldsymbol{B}$、10×6 型的矩阵 $\boldsymbol{C}$，且

$$\boldsymbol{B}=\begin{bmatrix}\boldsymbol{R}\\2\boldsymbol{R}\end{bmatrix},\boldsymbol{C}=\begin{bmatrix}\boldsymbol{R}&\boldsymbol{R}\\\boldsymbol{R}&\boldsymbol{0}\end{bmatrix}$$

其中 $\boldsymbol{R}$ 是行最简形矩阵，则 $\boldsymbol{B}$ 的秩是多少？$\boldsymbol{B}$ 的行阶梯型矩阵和 $\boldsymbol{C}$ 的行最简形是怎样的？若已知 $\boldsymbol{R}$ 的秩为 3，$\boldsymbol{N}(\boldsymbol{C}^{\boldsymbol{T}})$ 的维数是多少？
答：$\boldsymbol{B}$ 的秩等于 $\boldsymbol{R}$ 的秩；$\boldsymbol{B}$ 的行阶梯型矩阵和 $\boldsymbol{C}$ 的行最简形分别为

$$\begin{bmatrix}\boldsymbol{R}\\\boldsymbol{0}\end{bmatrix},\begin{bmatrix}\boldsymbol{R}&\boldsymbol{0}\\\boldsymbol{0}&\boldsymbol{R}\end{bmatrix}$$

(注：$\boldsymbol{C}$ 的行最简形还需要将零行交换到矩阵底部)；$\boldsymbol{N}(\boldsymbol{C}^{\boldsymbol{T}})$ 维数是 4

例四

考虑如下方程组及其解：

$$\boldsymbol{A}\boldsymbol{x}=\begin{bmatrix}2\\4\\2\end{bmatrix},\boldsymbol{x}=\begin{bmatrix}2\\0\\0\end{bmatrix}+c\begin{bmatrix}1\\1\\0\end{bmatrix}+d\begin{bmatrix}0\\0\\1\end{bmatrix}$$

$\boldsymbol{A}$ 行空间的维数？$\boldsymbol{A}$ 等于多少？$\boldsymbol{b}$ 满足什么条件才使得 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 有解？
答：一维；

$$\boldsymbol{A}=\begin{bmatrix}1&-1&0\\2&-2&0\\1&-1&0\end{bmatrix};\boldsymbol{b}=c\begin{bmatrix}1\\2\\1\end{bmatrix}$$

例五

方阵 $\boldsymbol{A}$ 的零空间只包含零向量，$\boldsymbol{A}^{\boldsymbol{T}}$ 的零空间？
答：只包含零向量

例六

所有的 5 阶可逆矩阵是否构成矩阵空间？
答：否 (不包括零矩阵)

例七

是否存在非零矩阵 $\boldsymbol{A}$ 满足 $\boldsymbol{A}^2=\boldsymbol{0}$？

答：存在，比如

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

例八

方阵的列向量线性无关，$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 是否总是有解？
答：是

例九

假设

$$\boldsymbol{B}=\boldsymbol{C}\boldsymbol{D}=\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & -1 & 2 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \boldsymbol{b}=\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

其中 $\boldsymbol{C}$ 是可逆矩阵，$\boldsymbol{D}$ 秩为 2，则 $\boldsymbol{N}(\boldsymbol{C}\boldsymbol{D})=\boldsymbol{N}(\boldsymbol{D})$，求 $\boldsymbol{B}\boldsymbol{x}=\boldsymbol{b}$ 的通解？
答：(注意使用第七讲中的零空间矩阵 $\boldsymbol{N}$)

$$\boldsymbol{x}=\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}+c\begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix}+d\begin{bmatrix} -2 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

例十

方阵的行空间是否等于列空间？
答：否

例十一

如果 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的四个子空间相同，则 $\boldsymbol{A}$ 是 $\boldsymbol{B}$ 的倍数
答：错误 (任意的同型的可逆矩阵的四个子空间都相同)

例十二

交换矩阵的两行，哪些子空间没变？
答：行空间和零空间

例十三

为什么向量 $(1,2,3)$ 不能既是矩阵 $\boldsymbol{A}$ 的行向量，又在 $\boldsymbol{A}$ 的零空间中？
答：行空间和零空间正交，两个空间的交集只有零向量，事实上对于任何非零向量都不可能既在行空间中，又在零空间中

第十四讲　正交向量与子空间

大纲

正交向量与子空间、零空间 ⊥ 行空间、$N(A^TA)=N(A)$

向量正交

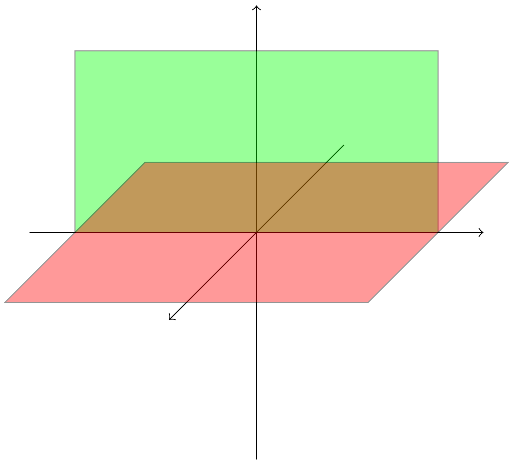
两个向量是正交的 (*orthogonal*) 是指它们之间是垂直的，即这两个向量的点积 (*doc product*) 等于 0。

向量空间正交

两个向量空间 A、B 是正交的 (*orthogonal*) 是指 A 中的任意一个向量和 B 中的任意一个向量都是正交的。

示例一

如下两个平面空间是否正交?



不正交，因为两个平面空间的交集有非零向量 (正交的向量空间的交集一定没有非零向量)

零空间 ⊥ 行空间

根据这两个空间的定义，很容易的可以得到零空间 ⊥ 行空间。并且这两个子空间的维数相加正好等于整个空间 $\mathbb{R}^n$ 的维数 n，所以称这两个空间是 $\mathbb{R}^n$ 中的正交补 (*orthogonal complements*)。(注：这里的补并不是说对于 $\mathbb{R}^n$ 中任意一个向量，它要么在行空间中，要么在零空间中，事实上存在无数多个向量既不在行空间中，也不在零空间中，比如三维空间中的一个 (过原点) 平面和垂直于该平面的 (过原点) 直线并没有充满整个 $\mathbb{R}^3$)

$Ax=b$ 的最优解

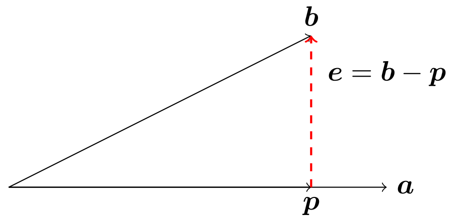
当 $Ax=b$ 无解时 (第八讲中的内容)，通过求解 $A^TA\hat{x}=A^Tb$ 来求解原方程的最优解。A^TA 有几个性质：

- A^TA 总是对称方阵
- A^TA 的秩与 A 的秩相等
- $N(A^TA)=N(A)$

第十五讲　子空间投影

投影

向量之间的投影：给定两个向量 $\boldsymbol{a}$、$\boldsymbol{b}$，表示向量 $\boldsymbol{b}$ 在 $\boldsymbol{a}$ 上的投影 $\boldsymbol{p}$



首先 $\boldsymbol{p}$ 等于某倍的 $\boldsymbol{a}$：$\boldsymbol{p}=x\boldsymbol{a}$，而且 $\boldsymbol{e}=\boldsymbol{b}-\boldsymbol{p}$ 正交于 $\boldsymbol{a}$，即 $\boldsymbol{a}^T\boldsymbol{e}=0$，所以 $\boldsymbol{a}^T\boldsymbol{e}=\boldsymbol{a}^T(\boldsymbol{b}-\boldsymbol{p})=\boldsymbol{a}^T(\boldsymbol{b}-x\boldsymbol{a})=0$，可得

$$x=\frac{\boldsymbol{a}^T\boldsymbol{b}}{\boldsymbol{a}^T\boldsymbol{a}},\quad \boldsymbol{p}=x\boldsymbol{a}=\boldsymbol{a}x=\boldsymbol{a}\frac{\boldsymbol{a}^T\boldsymbol{b}}{\boldsymbol{a}^T\boldsymbol{a}}$$

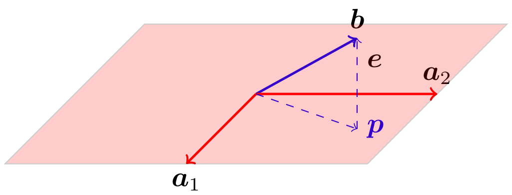
投影矩阵 (***Projection Matrix***)：

$$\boldsymbol{P}=\frac{\boldsymbol{a}\boldsymbol{a}^T}{\boldsymbol{a}^T\boldsymbol{a}}$$

则 $\boldsymbol{p}$ 可表示为 $\boldsymbol{p}=\boldsymbol{P}\boldsymbol{b}$。投影矩阵 $\boldsymbol{P}$ 有两个非常重要的性质：

- $\boldsymbol{P}^T=\boldsymbol{P}$
- $\boldsymbol{P}^2=\boldsymbol{P}$(几何空间的意义：投影两次和投影一次的效果相同)

向量在平面上的投影：$\boldsymbol{a}_1$、$\boldsymbol{a}_2$ 是平面的一组基向量，表示向量 $\boldsymbol{b}$ 在该平面上的投影 $\boldsymbol{p}$



$\boldsymbol{p}$ 在该平面内，所以可表示为 $\boldsymbol{a}_1$、$\boldsymbol{a}_2$ 的线性组合：$\boldsymbol{p}=\hat{x}_1\boldsymbol{a}_1+\hat{x}_2\boldsymbol{a}_2$，令 $\boldsymbol{A}=[\boldsymbol{a}_1,\boldsymbol{a}_2]$、$\hat{\boldsymbol{x}}=[\hat{x}_1,\hat{x}_2]^T$，则 $\boldsymbol{p}$ 可表示为 $\boldsymbol{p}=\boldsymbol{A}\hat{\boldsymbol{x}}$，$\boldsymbol{e}$ 可表示为 $\boldsymbol{e}=\boldsymbol{b}-\boldsymbol{p}=\boldsymbol{b}-\boldsymbol{A}\hat{\boldsymbol{x}}$，且 $\boldsymbol{e}$ 与 $\boldsymbol{a}_1$、$\boldsymbol{a}_2$ 均正交，即 $\boldsymbol{a}_1^T\boldsymbol{e}=0$、$\boldsymbol{a}_2^T\boldsymbol{e}=0$，将 $\boldsymbol{e}$ 代入可表示为

$$\boldsymbol{a}_1^T(\boldsymbol{b}-\boldsymbol{A}\hat{\boldsymbol{x}})=0,\quad \boldsymbol{a}_2^T(\boldsymbol{b}-\boldsymbol{A}\hat{\boldsymbol{x}})=0$$

表示成矩阵则为

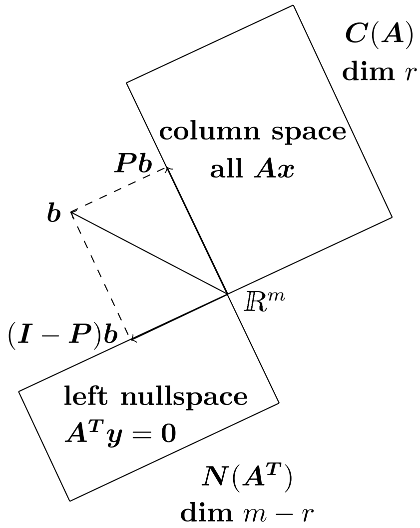
$$\begin{bmatrix}\boldsymbol{a}_1^T\\ \boldsymbol{a}_2^T\end{bmatrix}\begin{bmatrix}\boldsymbol{b}-\boldsymbol{A}\hat{\boldsymbol{x}}\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix},\quad \boldsymbol{A}^T(\boldsymbol{b}-\boldsymbol{A}\hat{\boldsymbol{x}})=\boldsymbol{0}$$

可解得 $\hat{\boldsymbol{x}}=(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b}$，则 $\boldsymbol{p}=\boldsymbol{A}\hat{\boldsymbol{x}}=\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b}$，投影矩阵为 $\boldsymbol{P}=\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$，同样满足上述两个性质：$\boldsymbol{P}^T=\boldsymbol{P}$、$\boldsymbol{P}^2=\boldsymbol{P}$。

第十六讲　投影矩阵和最小二乘

投影矩阵的作用

投影矩阵 $\boldsymbol{P}$ 的作用就是将 $\mathbb{R}^m$ 上的任意一个向量 $\boldsymbol{b}$ 投影到 $\boldsymbol{A}$ 的列空间上，对应的 $\boldsymbol{I}-\boldsymbol{P}$ 就是投影到左零空间上：



最小二乘法

用三个点 $(1,1)$、$(2,2)$、$(3,2)$ 来拟合直线方程。设直线方程为 $y=C+Dx$，则对应的方程组为

$$\begin{cases} C+D=1 \\ C+2D=2 \\ C+3D=2 \end{cases}$$

对应的矩阵形式 $\boldsymbol{Ax}=\boldsymbol{b}$ 为

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} C \\ D \end{bmatrix}=\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

$\boldsymbol{b}$ 并不在 $\boldsymbol{C}(\boldsymbol{A})$ 中（无解），所以将 $\boldsymbol{b}$ 投影到 $\boldsymbol{C}(\boldsymbol{A})$ 中再求解，则变为 $\boldsymbol{A}\hat{\boldsymbol{x}}=\boldsymbol{Pb}=\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b}$，两边左乘 $\boldsymbol{A}^T$，得 $\boldsymbol{A}^T\boldsymbol{A}\hat{\boldsymbol{x}}=\boldsymbol{A}^T\boldsymbol{b}$。

总结

这一讲以及上一讲都是在介绍投影矩阵及其应用，重点其实就是上面那张关于向量 $\boldsymbol{b}$ 是如何投影到列空间和左零空间上的图，当然还涉及到一些简单的证明，这里不作记录。

第十七讲　正交矩阵和格拉姆-施密特正交化

大纲

标准正交向量组、标准正交基、标准正交矩阵、格拉姆-施密特正交化

标准正交

如果一个向量组中的每一个向量的长度都是 1、并且彼此之间正交，则称该向量组是标准正交的 (*orthonormal*)；类似地，如果一个基是标准正交的，则称为标准正交基 (*orthonormal basis*)；如果一个 (实数) 方阵的行向量组和列向量组都是标准正交的，则称该矩阵为标准正交矩阵 (*orthonormal matrix*)。标准正交矩阵的一个特点在于它的转置矩阵等于逆矩阵，即 $\boldsymbol{Q}^T=\boldsymbol{Q}^{-1}, \boldsymbol{Q}^T\boldsymbol{Q}=\boldsymbol{I}$。

标准正交向量组或标准正交矩阵的这些特点会使得很多计算变得更加简单，比如投影矩阵 $\boldsymbol{P}=\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$，当矩阵 $\boldsymbol{A}$ 是正交矩阵时，上式可简化为 $\boldsymbol{P}=\boldsymbol{Q}\boldsymbol{Q}^T$；又比如求最优解时，$\boldsymbol{A}^T\boldsymbol{A}\hat{\boldsymbol{x}}=\boldsymbol{A}^T\boldsymbol{b}$ 则可直接简化为 $\hat{\boldsymbol{x}}=\boldsymbol{Q}^T\boldsymbol{b}$。

格拉姆-施密特

格拉姆-施密特正交化方法是将一组线性无关的向量或基向量进行标准正交化的方法。假设一组线性无关的向量分别为 $\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{v}_3,\cdots,\boldsymbol{v}_n$，则正交化过程为

$$
\begin{aligned}
\boldsymbol{u}_1 &= \boldsymbol{v}_1, & \boldsymbol{e}_1 &= \frac{\boldsymbol{u}_1}{||\boldsymbol{u}_1||}\\
\boldsymbol{u}_2 &= \boldsymbol{v}_2 - proj_{\boldsymbol{u}_1}(\boldsymbol{v}_2), & \boldsymbol{e}_2 &= \frac{\boldsymbol{u}_2}{||\boldsymbol{u}_2||}\\
\boldsymbol{u}_3 &= \boldsymbol{v}_3 - proj_{\boldsymbol{u}_1}(\boldsymbol{v}_3) - proj_{\boldsymbol{u}_2}(\boldsymbol{v}_3), & \boldsymbol{e}_3 &= \frac{\boldsymbol{u}_3}{||\boldsymbol{u}_3||}\\
&\cdots\\
\boldsymbol{u}_n &= \boldsymbol{v}_n - \sum_{i=1}^{n} proj_{\boldsymbol{u}_i}(\boldsymbol{v}_n), & \boldsymbol{e}_n &= \frac{\boldsymbol{u}_n}{||\boldsymbol{u}_n||}
\end{aligned}
$$

其中 $proj_{\boldsymbol{u}}(\boldsymbol{v})$ 为 $\boldsymbol{v}$ 在 $\boldsymbol{u}$ 上的投影 ($\langle\boldsymbol{u},\boldsymbol{v}\rangle$ 表示两者内积)

$$proj_{\boldsymbol{u}}(\boldsymbol{v}) = \frac{\langle\boldsymbol{u},\boldsymbol{v}\rangle}{\langle\boldsymbol{u},\boldsymbol{u}\rangle}\boldsymbol{u}$$

$\boldsymbol{A}=\boldsymbol{Q}\boldsymbol{R}$ 分解

而将上述格拉姆-施密特正交化过程用矩阵描述即为：$\boldsymbol{A}=\boldsymbol{Q}\boldsymbol{R}$，其中 $\boldsymbol{A}$ 的列向量是原始的线性无关的向量，$\boldsymbol{Q}$ 的列向量是标准正交化后的向量，$\boldsymbol{R}$ 是一个上三角矩阵 ($\boldsymbol{A}$ 中的任意一个列向量 $\boldsymbol{v}_j$ 都可以由 $\boldsymbol{Q}$ 的前 j 列 $\boldsymbol{e}_1$、$\boldsymbol{e}_2$、$\cdots$、$\boldsymbol{e}_j$ 线性组合出来)。(注：求解 $\boldsymbol{R}$ 有个小技巧，即两边左乘 $\boldsymbol{Q}^T$，则 $\boldsymbol{R}=\boldsymbol{Q}^T\boldsymbol{A}$)

总结

这一讲是关于正交的最后一讲，内容包括了标准正交的定义以及格拉姆-施密特正交化方法，注意从投影的角度去理解 (记忆为辅) 格拉姆-施密特正交化方法的过程。

第十八讲　行列式性质

行列式性质

行列式的十条性质：

1. $|\boldsymbol{I}| = 1$
2. 交换行列式的两行 (或列)，行列式的值取反
3. a. 行列式可以按行 (或列) 提出系数

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ kb_1 & kb_2 & kb_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = k\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}$$

   b. 如果行列式某行（或列）是两个因素之和，可把行列式拆成两个行列式之后相加

$$\begin{vmatrix} a_1+b_1 & a_2+b_2 & a_3+b_3 \\ c_1 & c_2 & c_3 \\ d_1 & d_2 & d_3 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 & a_3 \\ c_1 & c_2 & c_3 \\ d_1 & d_2 & d_3 \end{vmatrix} + \begin{vmatrix} b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \\ d_1 & d_2 & d_3 \end{vmatrix}$$

4. 如果行列式的两行 (或列) 相等，则行列式的值为 0
5. 行列式的一行 (或列) 加上另一行 (或列) 的任意倍，行列式的值不变

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1+ka_1 & b_2+ka_2 & b_3+ka_3 \\ c_1 & c_2 & c_3 \end{vmatrix}$$

6. 如果行列式中有一行 (或列) 均为 0，则行列式的值为 0
7. 上 (下) 三角矩阵对应的行列式的值等于对角线上的元素的积

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ & a_{22} & \cdots & a_{2n} \\ & & \ddots & \vdots \\ & & & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11} & & & \\ a_{21} & a_{22} & & \\ \vdots & \vdots & \ddots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = a_{11}a_{22}\cdots a_{nn}$$

8. $|\boldsymbol{A}| \neq 0 \Leftrightarrow \boldsymbol{A}$ 可逆
9. 若 $\boldsymbol{A}$、$\boldsymbol{B}$ 是 n 阶矩阵，$|\boldsymbol{AB}| = |\boldsymbol{A}| \cdot |\boldsymbol{B}|$；特别地，$|\boldsymbol{A}^2| = |\boldsymbol{A}|^2$
10. 若 $\boldsymbol{A}$ 是 n 阶矩阵，$|\boldsymbol{A}^T| = |\boldsymbol{A}|$

补充

这里再补充一些关于行列式的性质 (注：有些涉及到后面的概念)：

1. 若 $\boldsymbol{A}$ 是 n 阶矩阵，$|k\boldsymbol{A}| = k^n|\boldsymbol{A}|$
2. 若 $\boldsymbol{A}$ 是 n 阶可逆矩阵，$|\boldsymbol{A}^{-1}| = |\boldsymbol{A}|^{-1}$
3. 若 $\boldsymbol{A}$ 是 n 阶矩阵，$\boldsymbol{A}^*$ 是 $\boldsymbol{A}$ 的伴随矩阵，则 $|\boldsymbol{A}^*| = |\boldsymbol{A}|^{n-1}$

4. 若 $\boldsymbol{A}$ 的特征值是 $\lambda_i(i=1,2,\cdots,n)$，则 $|\boldsymbol{A}|=\prod_{i=1}^{n}\lambda_i$

5. 若矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 相似，即 $\boldsymbol{A}\sim\boldsymbol{B}$，则 $|\boldsymbol{A}|=|\boldsymbol{B}|$

6. 一般情况下，$|\boldsymbol{A}\pm\boldsymbol{B}|\neq|\boldsymbol{A}|\pm|\boldsymbol{B}|$

7. $|\boldsymbol{A}|\neq 0\Leftrightarrow\boldsymbol{A}$ 可逆 $\Leftrightarrow r(\boldsymbol{A})=n\Leftrightarrow\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}$ 没有非零解
$\Leftrightarrow 0$ 不是 $\boldsymbol{A}$ 的特征值 $\Leftrightarrow\boldsymbol{A}$ 的列 (行) 向量线性无关

8. 关于副对角线的行列式计算公式

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1,n-1} & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2,n-1} & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{n1} & 0 & \cdots & 0 & 0 \end{vmatrix} = \begin{vmatrix} 0 & \cdots & 0 & a_{1n} \\ 0 & \cdots & a_{2,n-1} & a_{2n} \\ \vdots & & \vdots & \vdots \\ a_{n1} & \cdots & a_{n,n-1} & a_{n,n} \end{vmatrix}$$
$$= (-1)^{\frac{n(n-1)}{2}} a_{1n}a_{2,n-1}\cdots a_{n1}$$

9. 两个特殊的拉普拉斯展开式 (如果 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 分别是 m 阶和 n 阶矩阵)

$$\begin{vmatrix} \boldsymbol{A} & * \\ \boldsymbol{0} & \boldsymbol{B} \end{vmatrix} = \begin{vmatrix} \boldsymbol{A} & \boldsymbol{0} \\ * & \boldsymbol{B} \end{vmatrix} = |\boldsymbol{A}|\cdot|\boldsymbol{B}|$$
$$\begin{vmatrix} \boldsymbol{O} & \boldsymbol{A} \\ \boldsymbol{B} & * \end{vmatrix} = \begin{vmatrix} * & \boldsymbol{A} \\ \boldsymbol{B} & \boldsymbol{O} \end{vmatrix} = |\boldsymbol{A}|\cdot|\boldsymbol{B}|\cdot(-1)^{mn}$$

10. 范德蒙行列式

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{vmatrix} = \prod_{1\leq j<i\leq n}(x_i-x_j)$$

总结

这一讲主要是关于行列式的基本性质。

第十九讲　行列式公式和代数余子式

行列式公式

先看一个 2 阶的行列式，先按第一行的元素分解，再按第二行元素分解 (上一讲的性质 3.b)：

$$\begin{aligned}\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} &= \begin{vmatrix} a_1 & 0 \\ b_1 & b_2 \end{vmatrix} + \begin{vmatrix} 0 & a_2 \\ b_1 & b_2 \end{vmatrix} \\ &= \begin{vmatrix} a_1 & 0 \\ b_1 & 0 \end{vmatrix} + \begin{vmatrix} a_1 & 0 \\ 0 & b_2 \end{vmatrix} + \begin{vmatrix} 0 & a_2 \\ b_1 & 0 \end{vmatrix} + \begin{vmatrix} 0 & a_2 \\ 0 & b_2 \end{vmatrix}\end{aligned}$$

对于 n 阶的行列式，则要分解 n 次，每次分解得到的行列式数量都是上一次分解的 n 倍，则一共可分解为 n^n 个行列式相加，其中每个行列式的每一行都选取了原行列式中对应行的一个元素 (其余为 0)，但这 n^n 个行列式中绝大部分都会有零列 (只要其中有一列的元素从没有被选中过)，只有 $n!$ 个行列式满足原行列的每一行、每一列都有元素被选中的条件。则行列式的公式 (也被称为 ***big formula***) 可表示为

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = \sum_{j_1 j_2 \cdots j_n} \pm a_{1j_1} a_{2j_2} \cdots a_{nj_n}$$

$j_1 j_2 \cdots j_n$ 表示 1 到 n 的某种排列，$\sum_{j_1 j_2 \cdots j_n}$ 表示对所有 $n!$ 种排列所对应的行列式求和。

(当然我们知道可以通过逆序数来确定前面的符号，更准确的表示为

$$\sum_{j_1 j_2 \cdots j_n} (-1)^{\tau(j_1 j_2 \cdots j_n)} a_{1j_1} a_{2j_2} \cdots a_{nj_n}$$

其中 $\tau(j_1 j_2 \cdots j_n)$ 表示排列 $j_1 j_2 \cdots j_n$ 的逆序数，即从左至右，看每个数后面比它小的数的数量，并将这些数量相加)

代数余子式

余子式 (***Minor***) 是指行列式 $|\boldsymbol{A}|$ 划去 a_{ij} 所在的第 i 行、第 j 列的元素，由原来的位置排法构成一个 $n-1$ 阶的行列式，记为 M_{ij}；代数余子式 (***Cofactor***) 记为 C_{ij}，$C_{ij} = (-1)^{i+j} M_{ij}$。(这两个翻译真是糟糕!)

按行（列）展开公式

$|\boldsymbol{A}|$ 按第 i 行展开的公式：

$$|\boldsymbol{A}| = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}, i = 1, 2, \cdots, n$$

$|\boldsymbol{A}|$ 按第 j 列展开的公式：

$$|\boldsymbol{A}| = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}, j = 1, 2, \cdots, n.$$

总结

这一讲和上一讲是介绍行列式的基本性质以及如何计算行列式，行列式的计算一般是通过行或列之间的加减 (性质 5) 来化简，使得行列式中包含尽可能多的 0 元素，再根据按行 (列) 展开公式来计算，一般不会直接分解计算 $n!$ 个行列式的和，除非行列式比较特殊。

第二十讲　克莱姆法则、逆矩阵、体积

大纲

代数余子式矩阵、伴随矩阵、逆矩阵、克莱姆法则、行列式的几何意义

代数余子式矩阵

代数余子式矩阵 (*Cofactor Matrix*) 即由代数余子式构成的矩阵，记为 $\boldsymbol{C}$：

$$\boldsymbol{C}=\begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{bmatrix}$$

伴随矩阵

伴随矩阵 (*Adjugate Matrix*) 记为 $\boldsymbol{A}^*$：

$$\boldsymbol{A}^*=\boldsymbol{C}^T=\begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix}$$

逆矩阵计算公式

$$\boldsymbol{A}^{-1}=\frac{1}{|\boldsymbol{A}|}\boldsymbol{A}^*$$

这个公式也很好推导：$\boldsymbol{A}\boldsymbol{A}^*=\boldsymbol{A}\boldsymbol{C}^T=|\boldsymbol{A}|\boldsymbol{I}$，这里有一个小技巧需要掌握：行列式的任一行（列）元素与另一行（列）元素的代数余子式乘积之和为 0 (性质 4)；而行列式的任一行（列）元素与它自身对应的代数余子式乘积之和为 $|\boldsymbol{A}|$ (也就是行列式按行或按列的展开公式)

克莱姆法则

克莱姆法则 (*Cramer's Rule*) 是用公式来求解 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 的方法，适用前提是 $\boldsymbol{A}$ 为可逆矩阵，则 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 可转化为

$$\boldsymbol{x}=\boldsymbol{A}^{-1}\boldsymbol{b}=\frac{1}{|\boldsymbol{A}|}\boldsymbol{A}^*\boldsymbol{b}=\frac{1}{|\boldsymbol{A}|}\boldsymbol{C}^T\boldsymbol{b}$$

而 $\boldsymbol{C}^T\boldsymbol{b}$ 会等于

$$\begin{aligned}\boldsymbol{C}^T\boldsymbol{b}&=\begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \\ &=\begin{bmatrix} b_1C_{11}+b_2C_{21}+\cdots+b_nC_{n1} \\ b_1C_{12}+b_2C_{22}+\cdots+b_nC_{n2} \\ \vdots \\ b_1C_{1n}+b_2C_{2n}+\cdots+b_nC_{nn} \end{bmatrix}=\begin{bmatrix} |\boldsymbol{B}_1| \\ |\boldsymbol{B}_2| \\ \vdots \\ |\boldsymbol{B}_n| \end{bmatrix}\end{aligned}$$

其中 $\boldsymbol{B}_i$ 为用 $\boldsymbol{b}$ 替换矩阵 $\boldsymbol{A}$ 中第 i 列所得的矩阵，故

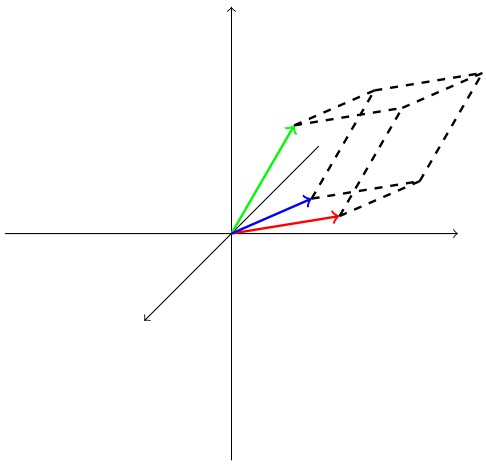
$$x_i = \frac{|\boldsymbol{B}_i|}{|\boldsymbol{A}|}$$

体积

三维行列式等价于几何空间中的体积；比如如下行列式

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

对应的空间图 (红绿蓝向量分别代表行向量，行列式的绝对值为体积，行列式的正负值代表左右手系，负值说明是左手系，正值说明右手系)：



第二十一讲　特征值和特征向量

特征值、特征向量

设 $\boldsymbol{A}$ 是 n 阶方阵，如果对于数 λ，存在非零向量 $\boldsymbol{x}$，使得 $\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{x}$，则 $\boldsymbol{A}$ 的特征值 (***Eigenvalues***) 为 λ，$\boldsymbol{A}$ 的特征向量 (***Eigenvalues***) 为 $\boldsymbol{x}(\boldsymbol{x}\neq\boldsymbol{0})$(特征向量有无限个，但线性无关的不是)。

特征向量的几何解释：矩阵 $\boldsymbol{A}$ 可视为一个变换 T_A，该变换的作用是将 $\mathbb{R}^n$ 中的一个向量变换为该空间的一个向量，而特征向量则是指在 T_A 变换前后，向量的方向相同、相反或者变换后为零向量的向量。

求解特征值和特征向量的算法：

1. 求解方程 $|\lambda\boldsymbol{I}-\boldsymbol{A}|=0$(在复数域内有 n 个根)，解即为特征值
2. 将特征值代入 $(\lambda\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}=\boldsymbol{0}$ 求解，解即为特征向量 (除零向量)

特征值的两个特点：

- $\sum_{i=1}^{n}\lambda_i=\sum_{i=1}^{n}a_{ii}$
- $\prod_{i=1}^{n}\lambda_i=|\boldsymbol{A}|$

示例一

投影矩阵 $\boldsymbol{P}=\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ 的特征值和特征向量有哪些特点？
答：对于在 $\boldsymbol{A}$ 列空间的向量 $\boldsymbol{x}$，$\boldsymbol{P}\boldsymbol{x}=\boldsymbol{x}$，特征值为 1；对于正交于 $\boldsymbol{A}$ 列空间的向量 $\boldsymbol{x}$ ($\boldsymbol{A}$ 左零空间中的向量)，$\boldsymbol{P}\boldsymbol{x}=\boldsymbol{0}$，特征值为 0

示例二

求如下矩阵的特征向量和特征值：

$$\boldsymbol{A}=\begin{bmatrix}3 & 1\\ 1 & 3\end{bmatrix}$$

$\boldsymbol{A}+3\boldsymbol{I}$ 的特征值和特征向量又是什么？
答：根据前述求解特征值和特征向量的算法，可求得 $\boldsymbol{A}$ 的特征值为：2 和 4，对应的特征向量为 $[1,1]^T$ 和 $[-1,1]^T$；$\boldsymbol{A}+3\boldsymbol{I}$ 的特征值为：$2+3$ 和 $4+3$，特征向量不变

示例三

求旋转矩阵 $\boldsymbol{Q}$ (该矩阵将向量旋转 $90°$) 的特征向量和特征值：

$$\boldsymbol{Q}=\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}$$

答：特征值为 (虚数)：i 和 $-i$

笔记

对称矩阵 ($\boldsymbol{A}^T=\boldsymbol{A}$) 的特征值为实数，反对称矩阵 ($\boldsymbol{A}^T=-\boldsymbol{A}$) 的特征值为虚数，这是两个极端

示例三

求如下三角矩阵的特征值和特征向量：

$$\boldsymbol{A} = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$$

答：特征值：$\lambda_1 = \lambda_2 = 3$，线性无关的特征向量只有 1 个 (小于 n，这时 $\boldsymbol{A}$ 无法相似对角化)，这就是特征向量的短缺情况

第二十二讲　对角化和矩阵的幂

大纲

矩阵对角化、矩阵的幂

对角化

假设 n 阶矩阵 $\boldsymbol{A}$ 有 n 个线性无关的特征向量，且

$$\boldsymbol{S}=\begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n\end{bmatrix},\boldsymbol{\Lambda}=\begin{bmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & \lambda_n\end{bmatrix}$$

$\boldsymbol{S}$ 和 $\boldsymbol{\Lambda}$ 分别是由特征向量和特征值构成的矩阵，则 $\boldsymbol{AS}=\boldsymbol{S\Lambda}$，即 $\boldsymbol{A}=\boldsymbol{S\Lambda S}^{-1}$，$\boldsymbol{A}^n=(\boldsymbol{S\Lambda S}^{-1})(\boldsymbol{S\Lambda S}^{-1})\cdots(\boldsymbol{S\Lambda S}^{-1})=\boldsymbol{S\Lambda^n S^{-1}}$

补充

这里补充一些关于特征值和特征向量的性质：

- $\lambda_1\neq\lambda_2$ 是 $\boldsymbol{A}$ 的特征值 $\Rightarrow$ $\boldsymbol{A}$ 的对应于 λ_1,λ_2 的特征向量 $\boldsymbol{x}_1,\boldsymbol{x}_2$ 线性无关
- n 阶矩阵 $\boldsymbol{A}$ 可相似对角化 $\Leftrightarrow$ $\boldsymbol{A}$ 有 n 个线性无关的特征向量 $\Leftrightarrow$ $\boldsymbol{A}$ 的每一个 r_i 重特征值对应的线性无关特征向量个数等于该特征值的重数 r_i
- r_i 重特征值对应的线性无关的特征向量的个数 $\leq r_i$

示例一

若矩阵 $\boldsymbol{A}$ 有 n 个线性无关的特征向量，$\boldsymbol{A}^k$ 什么时候趋于零矩阵?
答：所有特征值的绝对值均小于 1

示例二

下列矩阵是否可以对角化

$$\boldsymbol{A}=\begin{bmatrix}2 & 1\\ 0 & 2\end{bmatrix}$$

答：否 (没有 n 个线性无关的特征向量)

示例三

给定 (n 维) 向量 $\boldsymbol{u}_0$，且 $\boldsymbol{u}_{k+1}=\boldsymbol{A}\boldsymbol{u}_k$，求 $\boldsymbol{u}_k$ 的表达式 ($\boldsymbol{A}$ 可对角化)
答：$\boldsymbol{u}_0$ 可写成 $\boldsymbol{A}$ 的特征向量的线性组合形式 $c_1\boldsymbol{x}_1+c_2\boldsymbol{x}_2+\cdots+c_n\boldsymbol{x}_n$，令 $\boldsymbol{C}=[c_1,c_2,\cdots,c_n]^T$，则

$$\begin{aligned}\boldsymbol{u}_k=\boldsymbol{A}^k\boldsymbol{u}_0&=\boldsymbol{A}^k(c_1\boldsymbol{x}_1+c_2\boldsymbol{x}_2+\cdots+c_n\boldsymbol{x}_n)\\&=c_1\lambda_1^k\boldsymbol{x}_1+c_2\lambda_2^k\boldsymbol{x}_2+\cdots+c_n\lambda_n^k\boldsymbol{x}_n\\&=\boldsymbol{S\Lambda}^k\boldsymbol{C}\end{aligned}$$

示例四

斐波那契数列第 100 项是大致多少？增长速度多快？

答：$F_{k+2}=F_{k+1}+F_k$，令 $\boldsymbol{u}_k=[F_{k+1},F_k]^T$，则

$$\boldsymbol{u}_{k+1}=\begin{bmatrix}1&1\\1&0\end{bmatrix}\boldsymbol{u}_k=\boldsymbol{A}\boldsymbol{u}_k$$

矩阵 $\boldsymbol{A}$ 的特征值为 $\lambda_1=(1+\sqrt{5})/2\approx 1.618$ 和 $\lambda_2=(1-\sqrt{5})/2\approx -0.618$，其中 λ_1 绝对值大于 1，λ_2 的绝对值小于 1(可忽略)，则根据示例三可得

$$\begin{aligned}\boldsymbol{u}_{100}&=c_1\lambda_1^{100}\boldsymbol{x}_1+c_2\lambda_2^{100}\boldsymbol{x}_2\\&=c_1(\frac{1+\sqrt{5}}{2})^{100}\boldsymbol{x}_1+c_2(\frac{1-\sqrt{5}}{2})^{100}\boldsymbol{x}_2\\&\approx c_1(\frac{1+\sqrt{5}}{2})^{100}\boldsymbol{x}_1\end{aligned}$$

可以看出斐波那契数列的增长速率主要是由绝对值大于 1 的特征值决定的 (这里并未计算出特征向量和组合系数，因为并不影响该例子所想表达的思想)。

第二十三讲　微分方程和矩阵指数

矩阵指数

矩阵指数 (***Matrix Exponential***) 是和普通的指数函数类似的、作用于方阵的矩阵函数，主要用于解决线性微分方程。

$$e^{\boldsymbol{X}}=\sum_{k=0}^{\infty}\frac{1}{k!}\boldsymbol{X}^k$$

示例一

假设 $u_1=u_1(t)$ 和 $u_2=u_2(t)$，且满足 $u_1(0)=1$ 和 $u_2(0)=0$，求解如下微分方程组：

$$\frac{\mathrm{d}u_1}{\mathrm{d}t}=-u_1+2u_2,\ \frac{\mathrm{d}u_2}{\mathrm{d}t}=u_1-2u_2$$

答：令 $\boldsymbol{u}=[u_1,u_2]^T$，则方程组可表示为

$$\frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t}=\begin{bmatrix}\frac{\mathrm{d}u_1}{\mathrm{d}t}\\ \frac{\mathrm{d}u_2}{\mathrm{d}t}\end{bmatrix}=\begin{bmatrix}-1 & 2\\ 1 & -2\end{bmatrix}\begin{bmatrix}u_1\\ u_2\end{bmatrix}=\boldsymbol{A}\boldsymbol{u},\ \boldsymbol{A}=\begin{bmatrix}-1 & 2\\ 1 & -2\end{bmatrix}$$

求解 $\boldsymbol{A}$ 的特征值：$\lambda_1=0$、$\lambda_2=-3$；对应的特征向量为 $\boldsymbol{x}_1=[2,1]^T$、$\boldsymbol{x}_2=[1,-1]^T$。方程组的两个特解为：

$$\boldsymbol{w}_1=\begin{bmatrix}u_1(t)\\ u_2(t)\end{bmatrix}=e^{\lambda_1 t}\boldsymbol{x}_1,\ \boldsymbol{w}_2=\begin{bmatrix}u_1(t)\\ u_2(t)\end{bmatrix}=e^{\lambda_2 t}\boldsymbol{x}_2$$

简单地证明一下这两个为什么是原方程组的解：(只证 $\boldsymbol{w}_1$，$\boldsymbol{w}_2$ 类似)

$$\frac{\mathrm{d}\boldsymbol{w}_1}{\mathrm{d}t}=\frac{\mathrm{d}(e^{\lambda_1 t}\boldsymbol{x}_1)}{\mathrm{d}t}=\lambda_1 e^{\lambda_1 t}\boldsymbol{x}_1=e^{\lambda_1 t}\boldsymbol{A}\boldsymbol{x}_1=\boldsymbol{A}\boldsymbol{w}_1$$

方程组的通解为

$$\boldsymbol{u}(t)=c_1\boldsymbol{w}_1+c_2\boldsymbol{w}_2=c_1e^{0t}\begin{bmatrix}2\\ 1\end{bmatrix}+c_2e^{-3t}\begin{bmatrix}1\\ -1\end{bmatrix}$$

根据初始条件计算 c_1、c_2：$\boldsymbol{c}=\boldsymbol{S}^{-1}\boldsymbol{u}(0)$，其中 $\boldsymbol{c}=[c_1,c_2]$、$\boldsymbol{S}=[\boldsymbol{x}_1,\boldsymbol{x}_2]$

稳定性

并不是所有的系统都有稳定性，特征值可以告诉我们系统的状态：

1. 稳定性 (*Stability*)：当所有特征值的实数部分 <0 时，$\boldsymbol{u}(t)\to 0$
2. 稳态 (*Steady State*)：一个特征值为 0，其余所有特征值的实数部分均小于 0
3. 发散 (*Blow Up*)：存在特征值的实数部分 >0

示例一 (续)

使用特征向量矩阵 $\boldsymbol{S}$ 来解耦 (前提是 $\boldsymbol{A}$ 可对角化)：令 $\boldsymbol{v}=\boldsymbol{S}^{-1}\boldsymbol{u}$ 或 $\boldsymbol{u}=\boldsymbol{S}\boldsymbol{v}$，则原方程等价于

$$\frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t}=\boldsymbol{A}\boldsymbol{u} \Leftrightarrow \boldsymbol{S}\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t}=\boldsymbol{A}\boldsymbol{S}\boldsymbol{v} \Leftrightarrow \frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t}=\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}\boldsymbol{v}=\boldsymbol{\Lambda}\boldsymbol{v}$$

这个对角矩阵说明了方程之间的独立性 (解耦)：

$$\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t}=\boldsymbol{\Lambda}\boldsymbol{v} \Leftrightarrow \frac{\mathrm{d}v_i}{\mathrm{d}t}=\lambda_i v_i \Leftrightarrow v_i(t)=v_i(0)e^{\lambda_i t}$$

所以：

$$\boldsymbol{v}(t)=\begin{bmatrix} v_1(t) \\ v_2(t) \end{bmatrix}=\begin{bmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{bmatrix}\begin{bmatrix} v_1(0) \\ v_2(0) \end{bmatrix}=e^{\boldsymbol{\Lambda}t}\boldsymbol{v}(0)$$

$$\boldsymbol{u}(t)=\boldsymbol{S}\boldsymbol{v}(t)=\boldsymbol{S}e^{\boldsymbol{\Lambda}t}\boldsymbol{v}(0)=\boldsymbol{S}e^{\boldsymbol{\Lambda}t}\boldsymbol{S}^{-1}\boldsymbol{u}(0)=e^{\boldsymbol{A}t}\boldsymbol{u}(0)$$

为什么 $\boldsymbol{S}e^{\boldsymbol{\Lambda}t}\boldsymbol{S}^{-1}=e^{\boldsymbol{A}t}$?

$$\begin{aligned} e^{\boldsymbol{A}t} &= \boldsymbol{I}+\boldsymbol{A}t+\frac{(\boldsymbol{A}t)^2}{2!}+\cdots \\ &= \boldsymbol{S}\boldsymbol{S}^{-1}+\boldsymbol{S}\boldsymbol{\Lambda}\boldsymbol{S}^{-1}t+\frac{\boldsymbol{S}(\boldsymbol{\Lambda}t)^2\boldsymbol{S}^{-1}}{2!}+\cdots \\ &= \boldsymbol{S}(\boldsymbol{I}+\boldsymbol{\Lambda}t+\frac{(\boldsymbol{\Lambda}t)^2}{2!}+\cdots)\boldsymbol{S}^{-1} \\ &= \boldsymbol{S}e^{\boldsymbol{\Lambda}t}\boldsymbol{S}^{-1} \end{aligned}$$

$\boldsymbol{S}^{-1}\boldsymbol{u}(0)$ 是什么含义？前面已经说过，这个是通解中由初始条件确定的组合系数。所以可以推导出 $\boldsymbol{u}(t)$ 的通解为

$$\begin{aligned} \boldsymbol{u}(t) &= \boldsymbol{S}e^{\boldsymbol{\Lambda}t}\boldsymbol{S}^{-1}\boldsymbol{u}(0) \\ &= \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n \end{bmatrix}\begin{bmatrix} e^{\lambda_1 t} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2 t} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & e^{\lambda_n t} \end{bmatrix}\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \\ &= c_1\boldsymbol{x}_1e^{\lambda_1 t}+c_2\boldsymbol{x}_2e^{\lambda_2 t}+\cdots+c_n\boldsymbol{x}_ne^{\lambda_n t} \end{aligned}$$

示例二

求解二阶常微分方程：$y''+by'+ky=0, y=y(t)$

答：令 $\boldsymbol{u}=[y',y]^T$，则 $\boldsymbol{u}'=[y'',y']^T$，则原方程可表示为 (降阶)

$$\boldsymbol{u}'=\begin{bmatrix} y'' \\ y' \end{bmatrix}=\begin{bmatrix} -by'-ky \\ y \end{bmatrix}=\begin{bmatrix} -b & -k \\ 1 & 0 \end{bmatrix}\begin{bmatrix} y' \\ y \end{bmatrix}=\boldsymbol{A}\boldsymbol{u}$$

第二十四讲　马尔科夫矩阵；傅里叶级数

A^T 的特征值

A^T 的特征值和 A 的特征值相同，但特征向量不同。

马尔科夫矩阵

马尔科夫矩阵 (*Markov Matrix*) 是指所有的元素均为非负数且每一列的元素之和为 1 的方阵。马尔科夫矩阵具有如下特点：

1. 幂依旧是马尔科夫矩阵
2. $\lambda_1 = 1$ 是其一个特征值，其余特征值的绝对值均小于 1

示例一

在第二十二讲中讲矩阵的幂时介绍过一个示例 (示例三)：给定 (n 维) 向量 $\boldsymbol{u}_0$，且 $\boldsymbol{u}_{k+1} = \boldsymbol{A}\boldsymbol{u}_k$，求 $\boldsymbol{u}_k$ 的表达式 ($\boldsymbol{A}$ 为可对角化的 n 阶矩阵)，结果为 $\boldsymbol{u}_k = c_1\lambda_1^k\boldsymbol{x}_1 + c_2\lambda_2^k\boldsymbol{x}_2 + \cdots + c_n\lambda_n^k\boldsymbol{x}_n$，若 $\boldsymbol{A}$ 是马尔科夫矩阵，由上述性质可知 $\boldsymbol{u}_k$ 的稳态为 $c_1\boldsymbol{x}_1$。课程中介绍了一个加州和麻州之间人口迁移的例子，抽象之后就是上面的例子，其中

$$\boldsymbol{u}_0 = \begin{bmatrix} u_{Cal} \\ u_{Mass} \end{bmatrix}_0 = \begin{bmatrix} 0 \\ 1000 \end{bmatrix},\ \boldsymbol{A} = \begin{bmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{bmatrix}$$

可求得特征值 $\lambda_1 = 1$、$\lambda_2 = 0.7$，对应的特征向量为 $\boldsymbol{x}_1 = [2,1]^T$、$\boldsymbol{x}_2 = [1,-1]^T$，所以

$$\boldsymbol{u}_k = c_1 1^k \begin{bmatrix} 2 \\ 1 \end{bmatrix} + c_2(0.7)^k \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

根据 $\boldsymbol{u}_0$ 可以计算出 c_1 和 c_2

傅里叶级数

假设有一组标准正交基 $\boldsymbol{q}_1$、$\boldsymbol{q}_2$、$\cdots$、$\boldsymbol{q}_n$，则任意 n 维向量 $\boldsymbol{v}$ 都可以表示为 $\boldsymbol{v} = x_1\boldsymbol{q}_1 + x_2\boldsymbol{q}_2 + x_n\boldsymbol{q}_n$，且系数 x_i 等于 $\boldsymbol{q}_i^T\boldsymbol{v}$，用矩阵表示则为

$$\boldsymbol{v} = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots & \boldsymbol{q}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \boldsymbol{Q}\boldsymbol{x},\ \boldsymbol{x} = \boldsymbol{Q}^{-1}\boldsymbol{v} = \boldsymbol{Q}^T\boldsymbol{v}$$

类比标准正交基，看如下一组相互正交的函数：1、$\cos x$、$\sin x$、$\cos 2x$、$\sin 2x$、$\cdots$，这组函数基和之前的标准正交基向量组的区别有如下几点：

1. 以函数而非向量作为基，对应空间中的元素也是函数而非向量
2. 无限维的 (而非有限维)
3. 向量的正交是以点积 (内积) 为 0 来定义的，而函数的点积定义如下：

$$f^Tg = \int_0^{2\pi} f(x)g(x)\mathrm{d}x$$

则对于任意函数都可以表示为：$f(x) = a_0 + a_1\cos x + b_1\sin x + a_2\cos 2x + b_2\sin 2x + \cdots$，而要求解系数也可以类比向量的例子，比如要求解 $\cos x$ 的系数 a_1，只需要计算 $f(x)$ 与 $\cos x$ 的点积：

$$\int_0^{2\pi} f(x)\cos x\mathrm{d}x = \int_0^{2\pi} a_1\cos^2 x\mathrm{d}x = a_1\pi,\ a_1 = \frac{\int_0^{2\pi} f(x)\cos x\mathrm{d}x}{\pi}$$

第二十五讲　测验二 复习

测验二材料

- 正交矩阵 $\boldsymbol{Q}=[\boldsymbol{q}_1,\boldsymbol{q}_2,\cdots,\boldsymbol{q}_n]$。$\boldsymbol{Q}^T\boldsymbol{Q}=\boldsymbol{I}$。
  投影：求 $\boldsymbol{Ax}=\boldsymbol{b}$ 的最优解。
  使用格拉姆-施密特正交化方法从任意基计算出标准正交基。
- $\det \boldsymbol{A}$
  行列式性质。
  行列式的 *big formula* ($n!$ 项，注意每一项前面的符号)。
  代数余子式公式，代数余子式计算逆矩阵的公式。
- 特征值：$\boldsymbol{Ax}=\lambda\boldsymbol{x}$。
  $\det(\boldsymbol{A}-\lambda\boldsymbol{I})=0$。
  对角化：$\boldsymbol{S}^{-1}\boldsymbol{AS}=\boldsymbol{\Lambda}$ ($\boldsymbol{A}$ 有 n 个线性无关的特征向量)。
  矩阵的幂：$\boldsymbol{A}^k=(\boldsymbol{S\Lambda S}^{-1})^k=\boldsymbol{S\Lambda}^k\boldsymbol{S}^{-1}$。

示例一

如果 $\boldsymbol{a}=[2,1,1]^T$，则 $\boldsymbol{a}$ 的投影矩阵 $\boldsymbol{P}$ 是什么？$\boldsymbol{P}$ 的秩是多少？$\boldsymbol{P}$ 的列空间？$\boldsymbol{P}$ 的特征值？$\boldsymbol{P}$ 的特征值 1 所对应的特征向量？$\boldsymbol{u}_{k+1}=\boldsymbol{Pu}_k$，初始条件 $\boldsymbol{u}_0=[9,9,0]^T$ 时，$\boldsymbol{u}_k$ 等于多少？

示例二

给定三个数据点：$(1,4)$、$(2,5)$、$(3,8)$，找出过原点的最优解。

示例三

两个向量 $\boldsymbol{a}_1=[1,2,3]$ 和 $\boldsymbol{a}_2=[1,1,1]^T$ 确定了一个平面，求该平面内的两个正交向量 (格拉姆-施密特正交化)。

示例四

给定 4×4 型矩阵 $\boldsymbol{A}$，且有特征值 λ_1、λ_2、λ_3、λ_4。λ_i 满足什么条件使得 $\boldsymbol{A}$ 是可逆矩阵？$\det(\boldsymbol{A}^{-1})$ 等于多少？$(\boldsymbol{A}+\boldsymbol{I})$ 的迹是多少？

示例五

三对角线矩阵 (*tridiagonal matrices*)，比如

$$\boldsymbol{A}_4=\begin{bmatrix}1&1&0&0\\1&1&1&0\\0&1&1&1\\0&0&1&1\end{bmatrix}$$

令 $D_n=\det\boldsymbol{A}_n$。使用代数余子式计算 $D_n=aD_{n-1}+bD_{n-2}$ 中的 a 和 b。并利用计算出的 a 和 b 计算 D_n 的通式。

示例六

考虑满足如下规律的对称矩阵：

$$
\boldsymbol{A}_2=\begin{bmatrix}0&1\\1&0\end{bmatrix},\quad \boldsymbol{A}_3=\begin{bmatrix}0&1&0\\1&0&2\\0&2&0\end{bmatrix},\quad \boldsymbol{A}_4=\begin{bmatrix}0&1&0&0\\1&0&2&0\\0&2&0&3\\0&0&3&0\end{bmatrix},\quad \cdots
$$

求将向量投影至 $\boldsymbol{A}_3$ 的列空间的投影矩阵？$\boldsymbol{A}_3$ 的特征值和特征向量？将向量投影至 $\boldsymbol{A}_4$ 的列空间的投影矩阵？证明或证否：当 n 是奇数时，$\boldsymbol{A}_n$ 是奇异矩阵；当 n 是偶数时，$\boldsymbol{A}_n$ 是可逆矩阵。

第二十六讲　对称矩阵和正定性

对称矩阵

$\boldsymbol{A}=\boldsymbol{A}^T$。对于实对称矩阵，具有以下性质：

- 特征值全是实数
- 不同的特征值对应的特征向量相互正交
- 存在可逆矩阵 $\boldsymbol{S}$，使得 $\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}=\boldsymbol{\Lambda}$
- 存在正交矩阵 $\boldsymbol{Q}$，使得 $\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q}=\boldsymbol{Q}^T\boldsymbol{A}\boldsymbol{Q}=\boldsymbol{\Lambda}$，也可以写成 $\boldsymbol{A}=\boldsymbol{Q}^{-1}\boldsymbol{\Lambda}\boldsymbol{Q}=\boldsymbol{Q}^T\boldsymbol{\Lambda}\boldsymbol{Q}$ (任何具有 $\boldsymbol{Q}^T\boldsymbol{\Lambda}\boldsymbol{Q}$ 形式的矩阵本身就是对称的)，数学家将该性质称为谱定理 (***spectral theorem***)，并将特征值视为矩阵的"谱"；在力学上，也被称为主轴定理 (***principal axis theorem***)

为什么实对称矩阵的特征值是实数

$\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{x}$ 取共轭得 $\bar{\boldsymbol{A}}\bar{\boldsymbol{x}}=\bar{\lambda}\bar{\boldsymbol{x}}$，因为 $\boldsymbol{A}$ 是实数矩阵，所以 $\boldsymbol{A}\bar{\boldsymbol{x}}=\bar{\lambda}\bar{\boldsymbol{x}}$ (这证明了对于实数矩阵来说，特征值都是以共轭对的形式出现的)，再取转置得 $\bar{\boldsymbol{x}}^T\boldsymbol{A}^T=\bar{\boldsymbol{x}}^T\bar{\lambda}$，又因为 $\boldsymbol{A}$ 是对称的，所以 $\bar{\boldsymbol{x}}^T\boldsymbol{A}=\bar{\boldsymbol{x}}^T\bar{\lambda}$，再分别在两边右乘 $\boldsymbol{x}$ 得 $\bar{\boldsymbol{x}}^T\boldsymbol{A}\boldsymbol{x}=\bar{\boldsymbol{x}}^T\bar{\lambda}\boldsymbol{x}$，因为 $\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{x}$，所以等式左边等于 $\bar{\boldsymbol{x}}^T\lambda\boldsymbol{x}$，所以最后可得 $\bar{\boldsymbol{x}}^T\lambda\boldsymbol{x}=\bar{\boldsymbol{x}}^T\bar{\lambda}\boldsymbol{x}$，并且

$$\bar{\boldsymbol{x}}^T\boldsymbol{x}=\begin{bmatrix}\bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_n\end{bmatrix}\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}=|x_1|^2+|x_2|^2+\cdots+|x_n|^2$$

因为特征向量 $\boldsymbol{x}\neq\boldsymbol{0}$，所以 $\bar{\boldsymbol{x}}^T\boldsymbol{x}\neq 0$，所以 $\lambda=\bar{\lambda}$。

实对称矩阵 $\boldsymbol{A}$ 的特征值都是实数且特征向量正交，而如果 $\boldsymbol{A}$ 有复数元素，则当且仅当 $\boldsymbol{A}=\bar{\boldsymbol{A}}^T$ 时，$\boldsymbol{A}$ 的特征值都是实数、特征向量正交 (证明过程和上述过程类似)。

实对称矩阵是投影矩阵的组合

上面已经说过，实对称矩阵可以写成：

$$\begin{aligned}\boldsymbol{A}&=\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T\\&=\begin{bmatrix}\boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots & \boldsymbol{q}_n\end{bmatrix}\begin{bmatrix}\lambda_1 & 0 & \cdots & 0\\0 & \lambda_2 & \cdots & 0\\\vdots & \vdots & & \vdots\\0 & 0 & \cdots & \lambda_n\end{bmatrix}\begin{bmatrix}\boldsymbol{q}_1^T\\\boldsymbol{q}_2^T\\\vdots\\\boldsymbol{q}_n^T\end{bmatrix}\\&=\lambda_1\boldsymbol{q}_1\boldsymbol{q}_1^T+\lambda_2\boldsymbol{q}_2\boldsymbol{q}_2^T+\cdots+\lambda_n\boldsymbol{q}_n\boldsymbol{q}_n^T\end{aligned}$$

矩阵 $\boldsymbol{q}_k\boldsymbol{q}_k^T$ 是将向量投影至 $\boldsymbol{q}_k$ 上的投影矩阵，所以每个对称矩阵都是投影矩阵的组合。

特征值的正负

对称矩阵的正主元的个数等于正特征值的个数。(教材 6.4 节有证明)

正定矩阵

正定矩阵 (***positive definite matrix***) 是指特征值都大于 0 的对称矩阵。检查一个矩阵是否是正定矩阵的一个方法是检查是否所有的主元均大于 0。(正定矩阵的行列式一定大于 0，但行列式大于 0 的矩阵不一定是正定矩阵)

正定的充要条件

矩阵 $\boldsymbol{A}$ 是正定矩阵

$\Leftrightarrow \forall$ 非零向量 $\boldsymbol{x}$，$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} > 0$

$\Leftrightarrow \lambda_i > 0$ (若 $\exists\lambda_i = 0$，则 $\exists$ 非零向量 $\boldsymbol{x}$，$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} = 0$)

$\Leftrightarrow$ 所有的主元均 > 0

$\Leftrightarrow$ 全部顺序主子式 > 0

$\Leftrightarrow \boldsymbol{A} = \boldsymbol{D}^T\boldsymbol{D}$，$\boldsymbol{D}$ 是可逆矩阵

$\Leftrightarrow \boldsymbol{A}$ 合同于单位矩阵 $\boldsymbol{I}$

正定的必要条件

矩阵 $\boldsymbol{A}$ 是正定矩阵

$\Rightarrow$ 主对角元素 $a_{ii} > 0$

$\Rightarrow |\boldsymbol{A}| > 0$

第二十七讲　复数矩阵；快速傅里叶变换

复数向量

对于一个复数向量 $\boldsymbol{z}=[z_1,z_2,\cdots,z_n]^T\in\mathbb{C}^n$，定义其长度的平方为

$$|\boldsymbol{z}|^2=\bar{\boldsymbol{z}}^T\boldsymbol{z}=|z_1|^2+|z_2|^2+\cdots+|z_n|^2$$

记 $\boldsymbol{z}^H=\bar{\boldsymbol{z}}^T$，所以可以上式可简记为 $|\boldsymbol{z}|^2=\boldsymbol{z}^H\boldsymbol{z}$ (H 代表埃尔米特的名字，Hermite)；对于两个复数向量的内积定义为 $\boldsymbol{y}^H\boldsymbol{x}$

复数矩阵

对于实数矩阵来说，对称矩阵的定义是 $\boldsymbol{A}^T=\boldsymbol{A}$，而对于复数矩阵，一个更好的定义是 $\bar{\boldsymbol{A}}^T=\boldsymbol{A}$，这样的矩阵被称为埃尔米特矩阵 (***Hermitian Matrix***)，埃尔米特矩阵对角线上的元素必须是实数。和对称矩阵类似，埃尔米特矩阵的特征值也都是实数、特征向量正交。

和实数矩阵中的 (标准) 正交矩阵类似，同样可以定义复数矩阵中的“标准正交矩阵”(使用新的内积定义)，即行 (列) 向量的长度都为 1 且行 (列) 向量之间的内积为 0，这样的矩阵被称为酉矩阵 (***Unitary Matrics***)

离散傅里叶变换

傅里叶级数 (*Fourier series*) 是将周期性的函数或信号表示成一系列不同频率的函数之和的方法：

$$f(x)=a_0+a_1\cos x+b_1\sin x+a_2\cos 2x+b_2\sin 2x+\cdots$$

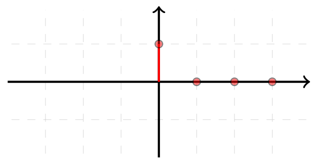
而要处理有限数据集时，离散傅里叶变换 (*Discrete Fourier Transform*，***DFT***) 则发挥至关重要的作用。这一讲主要就是讲如何用矩阵来表示离散傅里叶变换。在电子工程和计算机科学中，计数一般是从 0 开始到 $n-1$ 结束，而不是从 1 到 n，在这里也遵循这一传统来表示傅里叶矩阵 (矩阵的列向量都是正交的)：

$$\boldsymbol{F}_n=\begin{bmatrix}1&1&1&\cdots&1\\1&\omega&\omega^2&\cdots&\omega^{n-1}\\1&\omega^2&\omega^4&\cdots&\omega^{2(n-1)}\\\vdots&\vdots&\vdots&&\vdots\\1&\omega^{n-1}&\omega^{2(n-1)}&\cdots&\omega^{(n-1)^2}\end{bmatrix}$$

注意，$\boldsymbol{F}_n$ 不是酉矩阵，$\boldsymbol{F}_n/\sqrt{n}$ 才是；$\boldsymbol{F}_n$ 满足 $\boldsymbol{F}_n=\boldsymbol{F}_n^T$ 且 $(\boldsymbol{F}_n)_{jk}=\omega^{jk}$，其中 $j,k=0,1,\cdots,n-1$。这里的 $\omega=e^{i\cdot 2\pi/n},\omega^n=1$。
以 $\boldsymbol{F}_4$ 为例 (因为 $n=4$ 时，$\omega=e^{2\pi i/4}=i$，比较好表示)

$$\boldsymbol{F}_n=\begin{bmatrix}1&1&1&1\\1&i&i^2&i^3\\1&i^2&i^4&i^6\\1&i^3&i^6&i^9\end{bmatrix}=\begin{bmatrix}1&1&1&1\\1&i&-1&-i\\1&-1&1&-1\\1&-i&-1&i\end{bmatrix}$$

假设有一个在时刻 0 有单个脉冲的信号，如下图所示



可表示为 $[1,0,0,0]^T$，如果要计算该信号的离散傅里叶变换只需要左乘 $\boldsymbol{F}_4$ 得

$$\begin{bmatrix}1 & 1 & 1 & 1\\1 & i & -1 & -i\\1 & -1 & 1 & -1\\1 & -i & -1 & i\end{bmatrix}\begin{bmatrix}1\\0\\0\\0\end{bmatrix}=\begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$

可以看出该信号的离散傅里叶变换后所有的频率的幅度相等，如果再乘一次 $\boldsymbol{F}_4$ 又可以得到原始信号（注：4 倍是因为 $\boldsymbol{F}_4$ 不是酉矩阵，前面已经说过）

$$\begin{bmatrix}1 & 1 & 1 & 1\\1 & i & -1 & -i\\1 & -1 & 1 & -1\\1 & -i & -1 & i\end{bmatrix}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}=\boldsymbol{4}\begin{bmatrix}1\\0\\0\\0\end{bmatrix}$$

快速傅里叶变换

傅里叶可能并没有注意到傅里叶矩阵可以被分解成一些具有很多零元素的块，而高斯注意到了，但是他并没有意识到该发现是多么的重要！$\boldsymbol{F}_{2n}$ 和 $\boldsymbol{F}_n$ 之间有一个有趣的关系（注：$\omega_{2n}^2=\omega_n$）：

$$\boldsymbol{F}_{2n}=\begin{bmatrix}\boldsymbol{I} & \boldsymbol{D}\\\boldsymbol{I} & -\boldsymbol{D}\end{bmatrix}\begin{bmatrix}\boldsymbol{F}_n & \boldsymbol{0}\\\boldsymbol{0} & \boldsymbol{F}_n\end{bmatrix}\boldsymbol{P}$$

其中 $\boldsymbol{D}$ 是一个对角矩阵、$\boldsymbol{P}$ 是一个 $2n\times 2n$ 的置换矩阵（其作用是将向量的偶分量和奇分量（从 0 开始）分离出来，$\boldsymbol{Px}=[\boldsymbol{x}_{\text{even}},\boldsymbol{x}_{\text{odd}}]$）

$$\boldsymbol{D}=\begin{bmatrix}1 & 0 & 0 & \cdots & 0\\0 & \omega & 0 & \cdots & 0\\0 & 0 & \omega^2 & \cdots & 0\\\vdots & \vdots & \vdots & & \vdots\\0 & 0 & 0 & \cdots & \omega^{n-1}\end{bmatrix},\ \boldsymbol{P}=\begin{bmatrix}1 & 0 & 0 & 0 & \cdots & 0 & 0\\0 & 0 & 1 & 0 & \cdots & 0 & 0\\ & & & \vdots & & & \\0 & 0 & 0 & 0 & \cdots & 1 & 0\\0 & 1 & 0 & 0 & \cdots & 0 & 0\\0 & 0 & 0 & 1 & \cdots & 0 & 0\\ & & & \vdots & & & \\0 & 0 & 0 & 0 & \cdots & 0 & 1\end{bmatrix}$$

而对于 $\boldsymbol{F}_n$ 可以依次递归地分解下去，这就是快速傅里叶变换算法，所得到的结果是：对于 $\boldsymbol{F}_n$，如果直接计算需要 n^2 次操作，而快速傅里叶变换只需要 $\frac{1}{2}n\log_2 n$ 次操作，以 $n=1024$ 为例，前者需要上百万次操作，而后者只需要大约 5 千次操作。

示例一

给定一个 2×2 型的对称矩阵

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

第二十六讲的笔记中讲过，有一些方法可以判断它是否是正定矩阵：

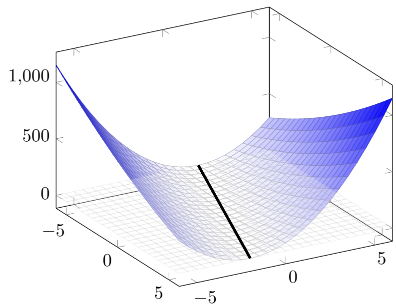
1. 检查特征值：$\lambda_1, \lambda_2 > 0$
2. 检查顺序主子式：$a > 0, ac - b^2 > 0$
3. 检查主元：$a > 0, (ac - b^2)/a > 0$
4. 对任意非零向量 $\boldsymbol{x}$，$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} > 0$

对于特定的 2×2 型矩阵

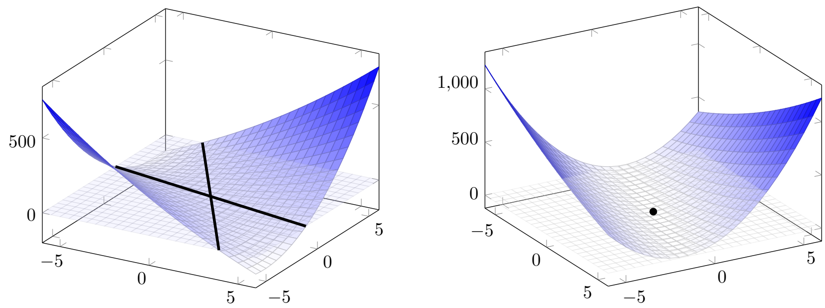
$$\boldsymbol{A} = \begin{bmatrix} 2 & 6 \\ 6 & y \end{bmatrix}$$

用行列式判断可知：当 $2y - 36 > 0$ 时（即 $y > 18$），$\boldsymbol{A}$ 是正定矩阵，$y = 18$ 是正定矩阵的边界；当正好 $y = 18$ 时，这样的矩阵称为半正定矩阵（***positive semidefinite matrix***），这是一个不可逆矩阵，特征值分别是 0 和 20，半正定矩阵的特征值大于或等于 0。

当 $y = 18$ 时，$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} = 2x_1^2 + 12x_1x_2 + 18x_2^2 = 2(x_1 + 3x_2)^2$，对应图像如下图所示，平面和曲面正好相切于 $x_1 = -3x_2$



而当 $y < 18$ 时（比如 $y = 7$ 时）和 $y > 18$ 时（比如 $y = 20$ 时），对应的图像如下图所示，可以看出前者曲面有一部分的值已经小于 0（在平面之下），而后者与平面的交点只有原点



主元、消元乘数与二次型系数的关系

示例一中 $y=20$ 时，

$$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}=2x_1^2+12x_1x_2+20x_2^2=\boxed{2}(x_1+\mathbf{3}x_2)^2+\boxed{2}x_2^2$$

$\boldsymbol{A}$ 的 $\boldsymbol{LU}$ 分解为

$$\boldsymbol{A}=\begin{bmatrix}2&6\\6&20\end{bmatrix}=\begin{bmatrix}1&0\\\mathbf{3}&1\end{bmatrix}\begin{bmatrix}\boxed{2}&6\\0&\boxed{2}\end{bmatrix}$$

可以看出主元是平方项的系数，消元乘数是括号中的内容。所以说对于一个正定矩阵来说，其主元为正数，所以 $\boldsymbol{x}$ 为非零向量时 $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$ 一定是大于 0 的。(注：有一种表述是“二次型的平方项都是正数，所以 $\boldsymbol{x}$ 为非零向量时 $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$ 一定是大于 0 的”，这样的表述正确的前提是可逆线性变换，比如 $\boldsymbol{LU}$ 分解中的 $\boldsymbol{L}$ 是可逆的)

补充 (二次型)

定义：n 个变量的二次型，系数均为实数时，称为 n 元实二次型，记为 $f(x_1,x_2,\cdots,x_n)$，其矩阵形式为

$$f(x_1,x_2,\cdots,x_n)=[x_1,x_2,\cdots,x_n]\begin{bmatrix}a_{11}&a_{11}&\cdots&a_{11}\\a_{21}&a_{22}&\cdots&a_{2n}\\\vdots&\vdots&&\vdots\\a_{n1}&a_{n2}&\cdots&a_{nn}\end{bmatrix}\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}=\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$$

其中 $\boldsymbol{A}$ 是对称矩阵

定理一：对任意一个 n 元二次型 $f(x_1,x_2,\cdots,x_n)=\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$，必存在正交变换 $\boldsymbol{x}=\boldsymbol{Q}\boldsymbol{y}$，化二次型为标准形，其中 $\boldsymbol{Q}$ 是正交矩阵

定理二：对任意一个 n 元二次型 $f(x_1,x_2,\cdots,x_n)=\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$，必存在可逆线性变换 $\boldsymbol{x}=\boldsymbol{C}\boldsymbol{y}$，化二次型为标准形，其中 $\boldsymbol{C}$ 是可逆矩阵

定理三：可逆线性变换不改变二次型的正定性

定理一等价表述：$\boldsymbol{A}$ 必既相似于又合同于对角阵，$\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q}=\boldsymbol{Q}^T\boldsymbol{A}\boldsymbol{Q}=\boldsymbol{\Lambda}$，正交变换不唯一，但得到的标准形唯一 (不考虑次序)

定理二等价表述：$\boldsymbol{A}$ 必合同于对角阵，$\boldsymbol{C}^T\boldsymbol{A}\boldsymbol{C}=\boldsymbol{\Lambda}$，可逆线性变换不唯一，标准形也不唯一，但规范形唯一（不考虑 $1,-1,0$ 的次序）

惯性定理：对于一个二次型，作可逆线性变换化成标准形（或规范形），可逆线性变换不唯一，标准形也不唯一，但其标准形中正平方项的项数 p，负平方项的项数 q 都是唯一确定的

海森矩阵

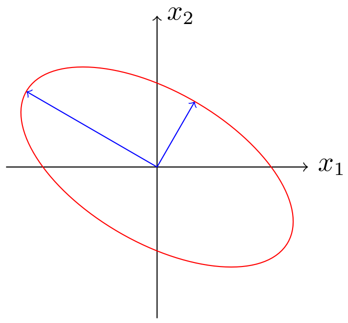
$f(x,y)$ 的二阶导数构成的矩阵：

$$\begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$$

因为 $f_{xy}=f_{yx}$，所以这是一个对称矩阵。当该矩阵为正定矩阵时，行列式 >0，也就是 $f_{xx}f_{yy}>f_{xy}^2$，这正好是微积分中计算最值的一个必要条件 (当导数存在时)

主轴定理

示例一中当 $y=20$ 时，$\boldsymbol{A}$ 是一个正定矩阵，$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}=2x_1^2+12x_1x_2+20x_2^2$，截取 $f(x_1,x_2)=1$ 可得到一个椭圆 (大概如下，不准确)，椭圆轴的方向是 $\boldsymbol{A}$ 特征向量的方向，而是椭圆轴的长度是 $1/\sqrt{\lambda_i}$。(而当 $y=7$ 时，$\boldsymbol{A}$ 不再是一个正定矩阵，截取 $f(x_1,x_2)=1$ 得到的是一个双曲线)



依次类推，当 $\boldsymbol{A}$ 为三维时，截取的就是一个椭球，轴的方向是 $\boldsymbol{A}$ 特征向量的方向，而是轴的长度是 $1/\sqrt{\lambda_i}$；n 维时，截取的就是一个 n 维的“椭球”等等

第二十九讲　相似矩阵和若尔当形

正定矩阵 (续)

如果 $\boldsymbol{A}$ 是对称的正定矩阵，它的逆是对称的正定矩阵吗?
答：是，$\boldsymbol{A}^{-1}$ 的特征值 $1/\lambda_1, 1/\lambda_2, \cdots, 1/\lambda_n$ 都大于 0
如果 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 都是正定的，$(\boldsymbol{A}+\boldsymbol{B})$ 是正定的吗?
答：是，利用正定的定义 (对任意的非零向量，$\boldsymbol{x}^T(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{x}$ 均大于 0)
当 $\boldsymbol{A}$ 是非方形的矩阵，在什么条件下 $\boldsymbol{A}^T\boldsymbol{A}$ 是正定的?
答：当 $\boldsymbol{A}$ 的列向量是线性无关的
正定矩阵的一个优点在于消元时不用进行行交换。

相似矩阵

设 $\boldsymbol{A}, \boldsymbol{B}$ 都是 n 阶方阵，若存在可逆矩阵 $\boldsymbol{M}$，使得 $\boldsymbol{M}^{-1}\boldsymbol{A}\boldsymbol{M} = \boldsymbol{B}$，则称 $\boldsymbol{A}$ 相似于 $\boldsymbol{B}$，记为 $\boldsymbol{A} \sim \boldsymbol{B}$。若 $\boldsymbol{A}$ 可对角化 $\boldsymbol{A} \sim \boldsymbol{\Lambda}$，则可以用 $\boldsymbol{\Lambda}$ 代表所有相似于 $\boldsymbol{\Lambda}$ 的矩阵所组成的家族。

相似矩阵的基本性质：

- 反身性：$\boldsymbol{A} \sim \boldsymbol{A}$
- 对称性：$\boldsymbol{A} \sim \boldsymbol{B} \Rightarrow \boldsymbol{B} \sim \boldsymbol{A}$
- 传递性：$\boldsymbol{A} \sim \boldsymbol{B}, \boldsymbol{B} \sim \boldsymbol{C} \Rightarrow \boldsymbol{A} \sim \boldsymbol{C}$

相似的必要条件：$\boldsymbol{A} \sim \boldsymbol{B} \Rightarrow$

- $|\lambda\boldsymbol{I} - \boldsymbol{A}| = |\lambda\boldsymbol{I} - \boldsymbol{B}|$ ($\Rightarrow \boldsymbol{A}, \boldsymbol{B}$ 的特征值相同)
- $\boldsymbol{A}, \boldsymbol{B}$ 的特征值相同
- $|\boldsymbol{A}| = |\boldsymbol{B}| = \sum_{i=1}^{n} \lambda_i$
- $\sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} b_{ii} = \sum_{i=1}^{n} \lambda_i$
- $\mathrm{r}(\boldsymbol{A}) = \mathrm{r}(\boldsymbol{B})$

相似矩阵的特征向量一般不同：假设 $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$，则

$$\begin{aligned}
\boldsymbol{A}\boldsymbol{M}\boldsymbol{M}^{-1}\boldsymbol{x} &= \lambda\boldsymbol{x} \\
\boldsymbol{M}^{-1}\boldsymbol{A}\boldsymbol{M}\boldsymbol{M}^{-1}\boldsymbol{x} &= \boldsymbol{M}^{-1}\lambda\boldsymbol{x} \\
\boldsymbol{B}\boldsymbol{M}^{-1}\boldsymbol{x} &= \boldsymbol{M}^{-1}\lambda\boldsymbol{x}
\end{aligned}$$

可知 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的特征值相同，但 $\boldsymbol{B}$ 的特征向量为 $\boldsymbol{M}^{-1}\boldsymbol{x}$

若尔当矩阵

若尔当矩阵 (***Jordan Matrix***)：除对角线和主对角线上方元素之外，其余元素均为 0，且对角线上方的元素若不为 0 只能为 1，且这 1 左方和下方的系数 (都在主对角线上) 有相同的值。
对角矩阵可以代表一个相似矩阵家族，但还有一些矩阵不能对角化 (没有 n 个线性无关的特征向量)，这时就可以用若尔当形来代表这些虽然相似但不可对角化的矩阵家族。(对角矩阵可以视为若尔当矩阵的特例，主对角线上方的元素都为 0)

若尔当定理

若尔当定理 (***Jordan's Theorem***) 是说每一个方阵 $\boldsymbol{A}$ 都相似于若尔当矩阵 $\boldsymbol{J}$，所有的若尔当块 (***Jordan Block***) 都在对角线上 (若尔当块是主对角线上都是相同的特征值、主对角线上方的元素都为 1、其余元素都为 0 的块)：

$$\boldsymbol{J}=\begin{bmatrix}\boldsymbol{J}_1 & 0 & \cdots & 0\\ 0 & \boldsymbol{J}_2 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & \boldsymbol{J}_d\end{bmatrix},\ \boldsymbol{J}_i=\begin{bmatrix}\lambda_i & 1 & 0 & \cdots & 0 & 0\\ 0 & \lambda_i & 1 & \cdots & 0 & 0\\ 0 & 0 & \lambda_i & \cdots & 0 & 0\\ \vdots & \vdots & \vdots & & & \vdots\\ 0 & 0 & 0 & \cdots & \lambda_i & 1\\ 0 & 0 & 0 & \cdots & 0 & \lambda_i\end{bmatrix}$$

若尔当块的个数 d 等于方阵 $\boldsymbol{A}$ 的线性无关的特征向量的个数，每一个若尔当块对应一个特征向量，而若尔当矩阵主对角线上方元素共有 $n-d$ 个 1。(若方阵 $\boldsymbol{A}$ 有 n 个线性无关的特征向量，则若尔当矩阵就变成了对角矩阵，主对角线上方的元素全都为 0)

示例一

$$\left[\begin{array}{c|c}4 & 0\\ \hline 0 & 4\end{array}\right],\ \begin{bmatrix}4 & 1\\ 0 & 4\end{bmatrix}$$

这两个若尔当矩阵分别代表了两个家族 (但特征值相同 $\lambda_1=\lambda_2=4$)，这两个家族之间的矩阵并不相似，只有家族内部的矩阵之间才相似。

示例二

$$\left[\begin{array}{ccc|c}0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0\\ \hline 0 & 0 & 0 & 0\end{array}\right],\ \left[\begin{array}{cc|cc}0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0\\ \hline 0 & 0 & 0 & 1\\ 0 & 0 & 0 & 0\end{array}\right]$$

这两个若尔当矩阵也分别代表了两个家族，虽然它们的特征值相同 (都为 0)，甚至线性无关的特征向量的个数也相同 (都为 2)，但这两个家族之间的矩阵依然并不相似 (只有家族内部的矩阵之间才相似)。

总结

这一讲介绍了用若尔当矩阵来表示一个相似矩阵组成的家族，不同的若尔当矩阵代表不同家族，这些家族之间的矩阵不是相似的，即使有相同的特征值；对角矩阵是若尔当矩阵的特例，对于非对角矩阵的若尔当矩阵，意味着线性无关的特征向量个数不够，也就是说有重特征值。

第三十讲　奇异值分解

奇异值分解

奇异值分解 (*singular value decomposition, SVD*) 是该课程中最后一个分解，也是最好的一个分解，表示为：

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$$

其中 $\boldsymbol{A}$ 是任意矩阵，$\boldsymbol{U}$ 是正交矩阵，$\boldsymbol{\Sigma}$ 是对角矩阵，$\boldsymbol{V}^T$ 是正交矩阵 ($\boldsymbol{SVD}$ 的一种特殊情况是当 $\boldsymbol{A}$ 是对称的正定矩阵时，$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T$)

关于 $\boldsymbol{SVD}$ 有两种形式，一种是标准的，一种是紧凑的，分别为

$$\boldsymbol{A}_{m\times n} = \boldsymbol{U}_{m\times m}\boldsymbol{\Sigma}_{m\times n}(\boldsymbol{V}^T)_{n\times n}$$
$$\boldsymbol{A}_{m\times n} = \boldsymbol{U}_{m\times r}\boldsymbol{\Sigma}_{r\times r}(\boldsymbol{V}^T)_{n\times r}$$

注意区分不同的上下文中所表达的意思。

工作原理：将 $\boldsymbol{A}$ 视为一个将 $\boldsymbol{C}(\boldsymbol{A}^T)$ 中的向量映射到 $\boldsymbol{C}(\boldsymbol{A})$ 的线性变换，而 $\boldsymbol{SVD}$ 就是用 $\boldsymbol{A}$ 将 $\boldsymbol{C}(\boldsymbol{A}^T)$ 的正交基变换成 $\boldsymbol{C}(\boldsymbol{A})$ 的正交基

$$\begin{aligned}\boldsymbol{A}\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_r\end{bmatrix} &= \begin{bmatrix}\sigma_1\boldsymbol{u}_1 & \sigma_2\boldsymbol{u}_2 & \cdots & \sigma_r\boldsymbol{u}_r\end{bmatrix}\\ &= \begin{bmatrix}\boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_r\end{bmatrix}\begin{bmatrix}\sigma_1 & 0 & \cdots & 0\\ 0 & \sigma_2 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & \sigma_r\end{bmatrix}\end{aligned}$$

其中 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_r$ 为 $\boldsymbol{C}(\boldsymbol{A}^T)$ 的一个标准正交基，$\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_r$ 为 $\boldsymbol{C}(\boldsymbol{A})$ 的一个标准正交基，$\sigma_1, \sigma_2, \cdots, \sigma_r$ 为伸缩因子。根据情况可以补全 $\boldsymbol{N}(\boldsymbol{A})$ 和 $\boldsymbol{N}(\boldsymbol{A}^T)$ 的标准正交基 $\boldsymbol{v}_{r+1}, \cdots, \boldsymbol{v}_n$ 和 $\boldsymbol{u}_{r+1}, \cdots, \boldsymbol{u}_m$，对应的伸缩因子为 0。

计算 $\boldsymbol{U}, \boldsymbol{\Sigma}, \boldsymbol{V}$ 的方法：$\boldsymbol{V}$ 的列向量是 $\boldsymbol{A}^T\boldsymbol{A}$ 的特征向量，$\boldsymbol{U}$ 的列向量是 $\boldsymbol{A}\boldsymbol{A}^T$ 的特征向量，而 $\boldsymbol{\Sigma}$ 可以由 $\boldsymbol{A}^T\boldsymbol{A}$ 或 $\boldsymbol{A}\boldsymbol{A}^T$ 的特征值决定

$$\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{V}\begin{bmatrix}\sigma_1^2 & 0 & \cdots & 0\\ 0 & \sigma_2^2 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & \sigma_n^2\end{bmatrix}\boldsymbol{V}^T,\ \boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{U}\begin{bmatrix}\sigma_1^2 & 0 & \cdots & 0\\ 0 & \sigma_2^2 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & \sigma_m^2\end{bmatrix}\boldsymbol{U}^T$$

示例一

求如下可逆矩阵的 $\boldsymbol{SVD}$ 分解：

$$\boldsymbol{A} = \begin{bmatrix}4 & 4\\ -3 & 3\end{bmatrix}$$

$$
\begin{aligned}
\boldsymbol{A}^T\boldsymbol{A} &= \begin{bmatrix} 4 & -3 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} \\
&= \begin{bmatrix} 25 & 7 \\ 7 & 25 \end{bmatrix}
\end{aligned}
$$

该矩阵的特征值为 32 和 18，对应的特征向量为 $\boldsymbol{v}_1 = [1/\sqrt{2}, 1/\sqrt{2}]^T$ 和 $\boldsymbol{v}_2 = [1/\sqrt{2}, -1/\sqrt{2}]^T$，所以

$$
\boldsymbol{\Sigma} = \begin{bmatrix} \sqrt{32} & 0 \\ 0 & \sqrt{18} \end{bmatrix}, \quad \boldsymbol{V}^T = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}
$$

再求 $\boldsymbol{U}$：

$$
\begin{aligned}
\boldsymbol{A}\boldsymbol{A}^T &= \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} \begin{bmatrix} 4 & -3 \\ 4 & 3 \end{bmatrix} \\
&= \begin{bmatrix} 32 & 0 \\ 0 & 18 \end{bmatrix}
\end{aligned}
$$

对应的特征向量为 $\boldsymbol{u}_1 = [1, 0]^T$ 和 $\boldsymbol{u}_2 = [0, 1]^T$，但因为 $\boldsymbol{A}\boldsymbol{v}_2 = [0, -\sqrt{18}]$，所以将 $\boldsymbol{u}_2$ 调整为 $\boldsymbol{u}_2 = [0, -1]^T$（这里要注意符号的调整），所以完整的分解为

$$
\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \sqrt{32} & 0 \\ 0 & \sqrt{18} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}
$$

示例二

求如下不可逆矩阵的 $\boldsymbol{SVD}$ 分解：

$$
\boldsymbol{A} = \begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix}
$$

行空间、列空间的标准基向量分别为 $\boldsymbol{v}_1 = [0.8, 0.6]^T$、$\boldsymbol{u}_1 = [1/\sqrt{5}, 2/\sqrt{5}]$，根据 $\boldsymbol{v}_2$、$\boldsymbol{u}_2$ 和 $\boldsymbol{v}_1$、$\boldsymbol{u}_1$ 的正交关系以及长度为 1 的限制，可求得 $\boldsymbol{v}_2 = [0.6, -0.8]^T$、$\boldsymbol{u}_2 = [2/\sqrt{5}, -1/\sqrt{5}]$，同时

$$
\begin{aligned}
\boldsymbol{A}\boldsymbol{A}^T &= \begin{bmatrix} 4 & 8 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix} \\
&= \begin{bmatrix} 80 & 60 \\ 60 & 45 \end{bmatrix}
\end{aligned}
$$

$\boldsymbol{A}\boldsymbol{A}^T$ 的一个特征值是 0，另一个是 125，所以完整的分解为

$$
\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} \sqrt{125} & 0 \\ 0 & \sqrt{0} \end{bmatrix} \begin{bmatrix} 0.8 & 0.6 \\ 0.6 & -0.8 \end{bmatrix}
$$

总结

$\boldsymbol{SVD}$ 分解涵盖了线性代数中的大部分重要内容，从四个子空间、特征值、特征向量到正定矩阵，并且具有非常广泛的应用。

第三十一讲　线性变换及其矩阵

线性变换

一个变换 T 是线性的是指对所有的向量 $\boldsymbol{v}, \boldsymbol{w}$ 和所有的标量 c 都满足：

- $T(\boldsymbol{v}+\boldsymbol{w}) = T(\boldsymbol{v}) + T(\boldsymbol{w})$
- $T(c\boldsymbol{v}) = cT(\boldsymbol{v})$

也可以等价的表示为：对 $\boldsymbol{v}, \boldsymbol{w}$ 和所有的标量 c, d 都满足 $T(c\boldsymbol{v}+d\boldsymbol{w}) = cT(\boldsymbol{v})+dT(\boldsymbol{w})$ (注意：$T(\boldsymbol{0})$ 必须等于 $\boldsymbol{0}$，否则不会满足 $T(c\boldsymbol{0}) = cT(\boldsymbol{0})$)

示例一

假设一个投影将 $\mathbb{R}^2$ 中的任意一个向量投影至 $\mathbb{R}^2$ 中过原点的一条直线上的另一个向量，我们可以将该投影看作是一个线性变换。

示例二

考虑一个进行平移的变换：$T(\boldsymbol{v}) = \boldsymbol{v} + \boldsymbol{v}_0$，其中 $\boldsymbol{v}_0$ 是一个非零常向量。这不是一个线性变换，因为 $T(2\boldsymbol{v}) = 2\boldsymbol{v} + \boldsymbol{v}_0 \neq 2T(\boldsymbol{v})$

示例三

$T(\boldsymbol{v}) = ||\boldsymbol{v}||$，该变换的结果是向量的长度。这不是一个线性变换，因为 $c < 0$ 时，$T(c\boldsymbol{v}) \neq cT(\boldsymbol{v})$

示例四

一个对向量进行逆时针旋转 $45°$ 的变换 T：$\mathbb{R}^2 \to \mathbb{R}^2$。这是一个线性变换。

用矩阵理解线性变换

通常来说，理解线性变换的最好的方法是找到变换背后隐藏的矩阵，从某种程度来说，线性变换只是对矩阵乘法的抽象描述。

示例五

定义 $T(\boldsymbol{v}) = \boldsymbol{A}\boldsymbol{v}$，其中 $\boldsymbol{A}$ 是一个矩阵。这是一个线性变换，因为满足：

- $\boldsymbol{A}(\boldsymbol{v}+\boldsymbol{w}) = \boldsymbol{A}(\boldsymbol{v}) + \boldsymbol{A}(\boldsymbol{w})$
- $\boldsymbol{A}(c\boldsymbol{v}) = c\boldsymbol{A}(\boldsymbol{v})$

如果 $\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$，从几何上来看应该如何描述这一变换?

答：关于 x 轴对称变换

示例六

如何找到一个将三维空间的向量映射到二维空间的向量的线性变换 T：$\mathbb{R}^3 \to \mathbb{R}^2$?

答：选择一个 2×3 的矩阵 $\boldsymbol{A}$，并定义 $T(\boldsymbol{v}) = \boldsymbol{A}\boldsymbol{v}$

描述 $T(\boldsymbol{v})$

需要知道哪些信息才能确定所有的 $\boldsymbol{v}$ 所对应的 $T(\boldsymbol{v})$?

答：输入空间的基 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$ 所对应的 $T(\boldsymbol{v}_1), T(\boldsymbol{v}_2), \cdots, T(\boldsymbol{v}_n)$。对于任意的 $\boldsymbol{v}$，可表示为 $\boldsymbol{v} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n$，则对应的 $T(\boldsymbol{v})$ 为 $T(\boldsymbol{v}) = c_1T(\boldsymbol{v}_1) + c_2T(\boldsymbol{v}_2) + \cdots + c_nT(\boldsymbol{v}_n)$

线性变换所对应的变换矩阵

给定一个线性变换 $T: \mathbb{R}^n \to \mathbb{R}^m$，如何构建一个对应的变换矩阵 $\boldsymbol{A}$ 来代表它？

答：分别选择输入空间和输出空间的一个基：$\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n$ 和 $\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_m$，满足

$$T(\boldsymbol{v}_i) = a_{1i}\boldsymbol{w}_1 + a_{2i}\boldsymbol{w}_2 + \cdots + a_{mi}\boldsymbol{w}_m$$

则可以将 $\boldsymbol{A}$ 表示为

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{n2} & \cdots & a_{mn} \end{bmatrix}$$

示例一 (续)

将 $\mathbb{R}^2$ 中的任意一个向量投影至 $\mathbb{R}^2$ 中过原点的一条直线上。此时 $m = n = 2$，为输入和输出空间选择相同的基，为了计算方便，其中一个基向量 $\boldsymbol{v}_1$ 选择为直线上的一个单位向量，$\boldsymbol{v}_2$ 选择为正交于 $\boldsymbol{v}_1$ 的单位向量，则这个投影变换 (在该基下) 所对应的矩阵为

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \boldsymbol{A}\boldsymbol{v} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} c_1 \\ 0 \end{bmatrix}$$

(Note：如果所选择的基正好是标准基下该投影变换所对应的矩阵 (即标准基下的投影矩阵) 的特征向量，则最后的变换矩阵 $\boldsymbol{A}$ 是一个对角矩阵，对角线上的元素为特征值)

为了明白基的选择的重要性，再看一个具体例子：$\mathbb{R}^2$ 中的任意一个向量投影至 $\mathbb{R}^2$ 中过原点、角度为 $45°$ 的一条直线上，但是基选择为标准基 $\boldsymbol{v}_1 = \boldsymbol{w}_1 = [1, 0]^T, \boldsymbol{v}_2 = \boldsymbol{w}_2 = [0, 1]^T$，对应的投影矩阵 (也就是标准基下的变换矩阵) 为

$$\boldsymbol{P} = \frac{\boldsymbol{a}\boldsymbol{a}^{\boldsymbol{T}}}{\boldsymbol{a}^{\boldsymbol{T}}\boldsymbol{a}} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

直接计算 $\boldsymbol{P}$ (标准基) 比前面计算 $\boldsymbol{A}$ (特征向量作为基) 更加困难。(注：$\boldsymbol{P}$ 和 $\boldsymbol{A}$ 只是同一变换在不同基下的描述方式；这里的标准基就不再是 $\boldsymbol{P}$ 的特征向量了，所以标准基下的变换矩阵 (也是 $\boldsymbol{P}$) 就不一定是对角矩阵了)。下一讲还会看到，对于同一变换、不同基下所对应的变换矩阵都是相似的，也就是说这里 $\boldsymbol{P}$ 和 $\boldsymbol{A}$ 是相似的。

示例七

令 T 是求导变换：$T(c_1+c_2x+c_3x^2)=c_2+2c_3$。输入空间的一个基是 $\boldsymbol{v}_1=1,\boldsymbol{v_2}=x,\boldsymbol{v_3}=x^2$，输出空间的一个基是 $\boldsymbol{w}_1=1,\boldsymbol{w}_2=x$，这也是一个线性变换。其中

$$\boldsymbol{A}=\begin{bmatrix}0&1&0\\0&0&2\end{bmatrix},\quad T(\begin{bmatrix}c_1\\c_2\\c_3\end{bmatrix})=\boldsymbol{A}\begin{bmatrix}c_1\\c_2\\c_3\end{bmatrix}=\begin{bmatrix}c_2\\2c_3\end{bmatrix}$$

总结

对于任意一个线性变换 T，都可找到一个矩阵 $\boldsymbol{A}$ 满足 $T(\boldsymbol{v})=\boldsymbol{A}\boldsymbol{v}$，如果变换是可逆的，则逆变换对应的矩阵为 $\boldsymbol{A}^{-1}$。两个变换 $T_1:\boldsymbol{v}\to\boldsymbol{A}_1\boldsymbol{v}$ 和 T_2：$\boldsymbol{w}\to\boldsymbol{A}_2\boldsymbol{w}$ 的乘积则对应于矩阵的乘积 $\boldsymbol{A}_1\boldsymbol{A}_2$。

第三十二讲　基的改变；图像压缩

图像压缩算法的基本思想

图像压缩算法的基本思想就是选择一个合适的基，以使得其中一小部分的基向量可以表示图像中的绝大部分信息，或者说使得其中一小部分的基向量比其它基向量更为重要，比如说只需要用 20% 的基向量就可以表示图像 99% 的信息，而剩余的 1% 的信息可能由于人眼分辨图像能力的有限性或者实际应用中可以容忍一小部分图像信息的损失等等原因，可以将另外的 80% 基向量全部忽略或忽略大部分。

示例一

举一个非常极端的例子，假设有一张 $m \times n$ 个像素点的黑白图像，每个像素点的取值范围是从 0 到 255，代表灰度强度，如果用一个向量来表示这样一张图像，这个向量是在 (mn) 维空间中的。以一张全黑图像为例，如果选择 mn 维空间中的标准基 (共 mn 个标准基向量) 来表示图像，这种情况下每个基向量都包含了相同数量的图像信息 (或者说它们是同等重要的)

$$\boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \cdots, \boldsymbol{e}_{(mn)} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

则这张图像可表示为 $255\boldsymbol{e}_1 + 255\boldsymbol{e}_2 + \cdots + 255\boldsymbol{e}_{(mn)}$，需要 $8mn$ bits 来存储 (每个 255 需要 8 bits)。而如果选择另外一组基，这组基中第一个基向量 $\boldsymbol{f}_1$ 为全 1 (比如傅里叶矩阵的第一列)，则该图像就可以表示为 $255\boldsymbol{f}_1$，只需要 8 bits 来存储

$$\boldsymbol{f}_1 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \cdots$$

可以看出，我们对基的选择会直接影响一张图像的大小或者一个视频的大小 (将视频的每一帧视为一张图像)，并且不同的场景下最优的基很有可能是不同的。

傅里叶基向量

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ \omega \\ \omega^2 \\ \omega^3 \\ \omega^4 \\ \omega^5 \\ \omega^6 \\ \omega^7 \end{bmatrix}, \begin{bmatrix} 1 \\ \omega^2 \\ \omega^4 \\ \omega^6 \\ \omega^8 \\ \omega^{10} \\ \omega^{12} \\ \omega^{14} \end{bmatrix}, \cdots, \begin{bmatrix} 1 \\ \omega^7 \\ \omega^{14} \\ \omega^{21} \\ \omega^{28} \\ \omega^{35} \\ \omega^{42} \\ \omega^{49} \end{bmatrix}$$

最出名的基就是傅里叶基，和之前学过的傅里叶矩阵有密切的关系。JPEG 使用的基是 ω^{jk} 的实数部分，也就是 cos 部分，也被称为离散余弦变换 (*Discrete Cosine Transform*，***DCT***)。JPEG 处理图像压缩的方法是将像素点分成 8×8 的块，每块共 64 个像素点。然后对每个块进行 ***DCT***，这一步是无损的，再设置人眼可分辨区别的阈值，忽略低于该阈值的系数 (注：JPEG 的图像压缩算法远比这里描述的复杂)

$$\text{signal } \boldsymbol{x} \xrightarrow{\text{lossless}} 64 \text{ coefficients } c \xrightarrow{\text{lossy compression}} \hat{c}(\text{many zero})$$

最后得到的图像为 $\hat{\boldsymbol{x}} = \sum \hat{c}_i \boldsymbol{v}_i$。在视频中，则不仅要考虑压缩每一帧，还要帧序列之间的关系，视频的每一帧和下一帧通常只会有很小的区别，所以只需要编码和压缩帧之间的区别，而不是完整的帧。

哈尔小波基

$$\begin{bmatrix}1\\1\\1\\1\\1\\1\\1\\1\end{bmatrix}, \begin{bmatrix}1\\1\\1\\1\\-1\\-1\\-1\\-1\end{bmatrix}, \begin{bmatrix}1\\1\\-1\\-1\\0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\0\\1\\1\\-1\\-1\end{bmatrix}, \begin{bmatrix}1\\-1\\0\\0\\0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\-1\\0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\0\\1\\-1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\0\\0\\0\\1\\-1\end{bmatrix}$$

在图像压缩算法中，小波基 (***wavelet basis***) 是傅里叶基的有力竞争者，上面是 $\mathbb{R}^8$ 中的小波基 (注：小波基也有很多的种类)，在原始的 JPEG (1997-2000) 中使用的是 ***DCT***，而在 JPEG2000 中使用的是离散小波变换 (*discrete wavelet transform*, ***DWT***)

基的要求

线性代数在图像压缩算法中发挥的作用主要是改变基 (比如上面的傅里叶基和小波基) 的情况下找出对应的系数 c_i，比如标准基表示的一个向量 x，用小波基可以表示成 (注：是同一个向量)

$$x = c_1\boldsymbol{w}_1 + c_2\boldsymbol{w}_2 + \cdots + c_8\boldsymbol{w}_8 = \boldsymbol{W}\boldsymbol{c}, \;\; \boldsymbol{c} = \boldsymbol{W}^{-1}\boldsymbol{x}$$

$\boldsymbol{c}$ 即为该向量在小波基下的表示。所以在图像压缩领域，一个好的基应该具有几个特点：

- 基矩阵的乘法和求逆非常快 (比如 ***FFT*** 或上面的小波基 $\boldsymbol{W}$)
- 好的压缩性-图像只需要很少的基向量就可以得到近似的图像，而大部分的基向量的系数都很小，可以直接忽略

基的改变和相似性

任何相似的变换矩阵本质上都是基的改变。考虑一个线性变换 $T: \mathbb{R}^n \to \mathbb{R}^n$，假设输入、输出空间使用的基 $\mathbb{R}^n$ 都是标准基，对应的变换矩阵为 $\boldsymbol{A}$，则

$$\mathbb{R}^n_{\text{standard}} \xrightarrow[\boldsymbol{A}]{T_{\boldsymbol{A}}} \mathbb{R}^n_{\text{standard}}$$

现在假设有 $\mathbb{R}^n$ 的另外一组基 $\boldsymbol{V}=[\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_n]$，对于用该基表示的任意一个向量 $\boldsymbol{x}$，左乘 $\boldsymbol{V}$ 所得的结果 $\boldsymbol{Vx}$ 就是该向量在标准基下的表示方法，或者说左乘 $\boldsymbol{V}$ 等价于将同一个向量在 $\boldsymbol{V}$ 基下的表示翻译成标准基下的表示，类似的，左乘 $\boldsymbol{V}^{-1}$ 正好相反，将同一个向量在标准基下的表示翻译成 $\boldsymbol{V}$ 基下的表示。所以对于同一个变换 T，如果是用 $\boldsymbol{V}$ 基来表示，可以翻译成标准基，再进行变换，再翻译回 $\boldsymbol{V}$ 基，所以可表示为 $\boldsymbol{B}=\boldsymbol{V}^{-1}\boldsymbol{AV}$，其中 $\boldsymbol{A}\sim\boldsymbol{B}$

$$\begin{tikzcd}
 & \mathbb{R}^n_{\text{standard}} \arrow[r, "T_{\boldsymbol{A}}", "\boldsymbol{A}"'] & \mathbb{R}^n_{\text{standard}} \arrow[dr, "\boldsymbol{V}^{-1}"] & \\
\mathbb{R}^n_{\{\boldsymbol{v}_i\}} \arrow[ur, "\boldsymbol{V}"] \arrow[rrr, "T_{\boldsymbol{A}}", "\boldsymbol{B}=\boldsymbol{V}^{-1}\boldsymbol{AV}"'] & & & \mathbb{R}^n_{\{\boldsymbol{v}_i\}}
\end{tikzcd}$$

对于其它的基也有类似的关系，比如 $\boldsymbol{W}=[\boldsymbol{w}_1,\boldsymbol{w}_2,\cdots,\boldsymbol{w}_n]$，对应的变换矩阵 $\boldsymbol{C}=\boldsymbol{W}^{-1}\boldsymbol{AW}$，同一变换的所有变换矩阵都是相似的 (不管基是什么)。而如果选择的基向量正好是对应的变换矩阵的特征向量，则变换矩阵为对角矩阵，这样的基向量非常的适合处理图像压缩，但找出这样的基相比找出傅里叶基或小波基需要更多的计算。

总结

这一讲简单地介绍了 (非常简单!!!) 图像压缩的基本思想、图像压缩常用的基以及基的改变和相似性的关系。总结这一讲中的思想：1. 图像压缩算法中和线性代数相关的内容主要是利用基的改变；2. 不同场景下好的基可能是不同的；3. 在不同基下的同一线性变换所对应的变换矩阵是相似的。如果想深入学习图像压缩，这一讲的内容可能只是一个导论，还需要专门去学习。

第三十三讲　测验三复习

测验三复习材料

- 特征值和特征向量
- 微分方程 $\mathrm{d}\boldsymbol{u}/\mathrm{d}t = \boldsymbol{A}\boldsymbol{u}$ 和矩阵指数 $e^{\boldsymbol{A}t}$
- 实对称矩阵
- 正定矩阵
- 相似矩阵
- $\boldsymbol{SVD}$

示例一

微分方程：

$$\frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t} = \boldsymbol{A}\boldsymbol{u} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \boldsymbol{u}$$

其通解形式为：$\boldsymbol{u}(t) = c_1 e^{\lambda_1 t}\boldsymbol{x}_1 + c_2 e^{\lambda_2 t}\boldsymbol{x}_2 + c_3 e^{\lambda_3 t}\boldsymbol{x}_3$

问：$\boldsymbol{A}$ 的特征值是多少？通解的周期是多少？说明 $\boldsymbol{A}$ 的两个特征向量是正交的。微分方程的解是 $\boldsymbol{u}(t) = e^{\boldsymbol{A}t}$，如何计算 $e^{\boldsymbol{A}t}$？

示例二

一个 3×3 的矩阵 $\boldsymbol{A}$ 有特征值 $\lambda_1 = 0, \lambda_2 = c, \lambda_3 = 2$，并且特征向量为

$$\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \boldsymbol{x}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \boldsymbol{x}_3 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$$

问：c 满足什么条件 $\boldsymbol{A}$ 是可对角化的？对称的？正定的？$\boldsymbol{A}$ 是马尔科夫矩阵吗？$\boldsymbol{P} = \frac{1}{2}\boldsymbol{A}$ 可以是投影矩阵吗？

示例三

$\boldsymbol{A}$ 的 $\boldsymbol{SVD}$ 分解 $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$，其中 $\boldsymbol{U}, \boldsymbol{\Sigma}, \boldsymbol{v}$ 都是 2×2 的，当 $\boldsymbol{\Sigma}$ 分别为

$$\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 3 & 0 \\ 0 & -5 \end{bmatrix}, \begin{bmatrix} 3 & 0 \\ 0 & 0 \end{bmatrix}$$

$\boldsymbol{A}$ 有什么特点？

示例四

$\boldsymbol{A}$ 是对称的、正交的。

问：$\boldsymbol{A}$ 的特征值有什么特点？$\boldsymbol{A}$ 一定是正定的吗？$\boldsymbol{A}$ 一定没有重特征值吗？$\boldsymbol{A}$ 可对角化吗？$\boldsymbol{A}$ 是可逆的吗？证明 $\boldsymbol{P} = \frac{1}{2}(\boldsymbol{A} + \boldsymbol{I})$ 是一个投影矩阵。

第三十四讲　左、右逆；伪逆

逆矩阵

之前学习的逆针对满秩矩阵 $(m=n=r)$，$\boldsymbol{A}^{-1}\boldsymbol{A}=\boldsymbol{A}\boldsymbol{A}^{-1}=\boldsymbol{I}$

左逆矩阵

当列满秩但行不满秩时 $(n=r<m)$，$\boldsymbol{A}^T\boldsymbol{A}$ 是一个可逆矩阵 (这是我们前面讨论最小二乘法时的核心)，即 $(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{A}=\boldsymbol{I}$，将 $\boldsymbol{A}$ 的左逆 (***left inverse***) 定义为 $\boldsymbol{A}_{\text{left}}^{-1}=(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$，$\boldsymbol{A}_{\text{left}}^{-1}\boldsymbol{A}=\boldsymbol{I}$(注：右乘左逆 $\boldsymbol{A}\boldsymbol{A}_{\text{left}}^{-1}$ 不是单位矩阵)

右逆矩阵

当行满秩但列不满秩时 $(m=r<n)$，类比左逆的定义，将 $\boldsymbol{A}$ 的右逆 (***right inverse***) 定义为 $\boldsymbol{A}_{\text{right}}^{-1}=\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}$，$\boldsymbol{A}\boldsymbol{A}_{\text{right}}^{-1}=\boldsymbol{I}$

左、右逆的个数

逆矩阵是唯一的，但对于左、右逆来说是有无数多个的，这里定义的是我们最喜欢的、形式上最好看的

$\boldsymbol{A}\boldsymbol{A}_{\text{left}}^{-1}$ 和 $\boldsymbol{A}_{\text{right}}^{-1}\boldsymbol{A}$

$\boldsymbol{A}\boldsymbol{A}_{\text{left}}^{-1}$ 是将 $\mathbb{R}^m$ 中的任一向量投影至 $\boldsymbol{A}$ 列空间的投影矩阵，$\boldsymbol{A}_{\text{right}}^{-1}\boldsymbol{A}$ 是将 $\mathbb{R}^n$ 中的任一向量投影至 $\boldsymbol{A}$ 行空间的投影矩阵

伪逆

当列不满秩且行不满秩时 $(r<m,r<n)$，对于 $\boldsymbol{A}$ 的行空间的任一向量 $\boldsymbol{x}$，左乘 $\boldsymbol{A}$ 所得结果 $\boldsymbol{A}\boldsymbol{x}$ 是 $\boldsymbol{A}$ 的列空间中的一个向量，并且是一一对应的，即如果 $\boldsymbol{x}\neq\boldsymbol{y}$，那么 $\boldsymbol{A}\boldsymbol{x}\neq\boldsymbol{A}\boldsymbol{y}$。$\boldsymbol{A}$ 的伪逆 (***pseudoinverse***) 记为 $\boldsymbol{A}^+$，满足对于 $\boldsymbol{A}$ 的行空间的任一向量 $\boldsymbol{x}$，$\boldsymbol{x}=\boldsymbol{A}^+\boldsymbol{A}\boldsymbol{x}$。$\boldsymbol{A}^+$ 的零空间就是 $\boldsymbol{A}^T$ 的零空间。

计算伪逆 $\boldsymbol{A}^+$：

1. 对 $\boldsymbol{A}$ 进行 *SVD* 分解：$\boldsymbol{A}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$，$\boldsymbol{\Sigma}$ 前 r 个对角线上的元素分别为 $\sigma_1,\sigma_2,\cdots,\sigma_r$，$\boldsymbol{U},\boldsymbol{V}$ 是标准正交矩阵
2. 求 $\boldsymbol{\Sigma}^+$：$\boldsymbol{\Sigma}^+$ 对角线上前 r 个的元素分别为 $1/\sigma_1,1/\sigma_2,\cdots,1/\sigma_r$
   $\boldsymbol{\Sigma}^+$ 满足：$\boldsymbol{\Sigma}^+$ 和 $\boldsymbol{\Sigma}$ 的乘积的对角线上的前 r 个元素都是 1
   $\boldsymbol{\Sigma}$ 是 $m\times n$ 型，$\boldsymbol{\Sigma}^+$ 是 $n\times m$ 型
3. $\boldsymbol{A}^+=\boldsymbol{V}\boldsymbol{\Sigma}^+\boldsymbol{U}^T$

第三十五讲　期末复习

示例一

$\boldsymbol{A}_{m\times n}$ 秩为 r，$\boldsymbol{A}\boldsymbol{x}=[1,0,0]^T$ 无解，$\boldsymbol{A}\boldsymbol{x}=[0,1,0]^T$ 有唯一解。
问：m,n,r 的值？给出一个满足条件的 $\boldsymbol{A}$。$|\boldsymbol{A}^T\boldsymbol{A}|$ 和 $|\boldsymbol{A}\boldsymbol{A}^T|$ 相等吗？$\boldsymbol{A}^T\boldsymbol{A}$ 可逆吗？$\boldsymbol{A}^T\boldsymbol{A}$ 正定吗？证明：对于任意 $\boldsymbol{c}$，$\boldsymbol{A}^T\boldsymbol{y}=\boldsymbol{c}$ 至少有一解。

示例二

$\boldsymbol{A}=[\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{v}_3]$。
求解 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{v}_1-\boldsymbol{v}_2+\boldsymbol{v}_3$。如果 $\boldsymbol{v}_1-\boldsymbol{v}_2+\boldsymbol{v}_3=\boldsymbol{0}$，$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{v}_1-\boldsymbol{v}_2+\boldsymbol{v}_3$ 的解唯一吗？如果 $\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{v}_3$ 是标准正交的，$\boldsymbol{v}_1,\boldsymbol{v}_2$ 的线性组合中哪个向量最接近 $\boldsymbol{v}_3$？

示例三

马尔科夫矩阵：

$$\boldsymbol{A}=\begin{bmatrix}0.2&0.4&0.3\\0.4&0.2&0.3\\0.4&0.4&0.4\end{bmatrix}$$

注：前两列的和是第三列的 2 倍。
问：$\boldsymbol{A}$ 的特征值是多少？$\boldsymbol{u}_k=\boldsymbol{A}^k\boldsymbol{u}(0),\boldsymbol{u}(0)=[0,10,0]^T$，则 $\lim_{k\to\infty}\boldsymbol{u}_k$ 是什么？

示例四

求分别满足以下条件的 2×2 型矩阵：

1. 投影至 $\boldsymbol{a}=[4,-3]^T$ 上的投影矩阵
2. 特征值为 $\lambda_1=0,\lambda_2=3$，特征向量为 $\boldsymbol{x}_1=[1,2]^T,\boldsymbol{x}_2=[2,1]^T$
3. 实矩阵且不能分解成 $\boldsymbol{B}^T\boldsymbol{B}$
4. 非对称，但有正交的特征向量

示例五

最小二乘法求最优解：

$$\boldsymbol{A}\begin{bmatrix}c\\d\end{bmatrix}=\begin{bmatrix}1&0\\1&1\\1&2\end{bmatrix}\begin{bmatrix}c\\d\end{bmatrix}=\begin{bmatrix}3\\4\\1\end{bmatrix}=\boldsymbol{b}$$

的最优解为 $[\hat{c},\hat{d}]^T=[11/3,-1]^T$。
问：$\boldsymbol{b}$ 在 $\boldsymbol{A}$ 的列空间中的投影 $\boldsymbol{p}$ 是什么？画出对应的直线。找出一个不同的 $\boldsymbol{b}\neq\boldsymbol{0}\in\mathbb{R}^3$，满足最优解为 $[\hat{c},\hat{d}]^T=[0,0]^T$